Tight exponential evaluation for universal composablity with privacy amplification and its applications

Masahito Hayashi

***Abstract*—We adopt the universal composablity as Eve's distinguishability in secret key generation from a common random number between two distinct players without communication. Under this secrecy criterion, using the Rényi entropy of order $1+s$ for $s \in [0,1]$, we derive a new upper bound of Eve's distinguishability under the application of the universal$_2$ hash functions. It is also shown that this bound gives the tight exponential rate of decrease in the case of independent and identical distributions. The result is applied to the wire-tap channel model and to secret key generation (distillation) by public discussion.**

***Index Terms*—sacrifice bits, L_1 norm distance, universal composablity, secret key distillation, universal$_2$ hash functions, wire-tap channel**

I. Introduction

Random privacy amplification based on the universal$_2$ condition[1] has been studied by many authors[2], [3], [4], [5], [30], [6]. This technique is originally aimed for random number extraction[2], [3]. It can be applied to secret key generation (distillation) with public communication[7], [8], [9], [10], [11], [3], [4] and wire-tap channel[12], [13], [14], [15], [16], [17], which treats the secure communication in the presence of an eavesdropper. (For details of its application, see e.g. the previous paper [6].) When random privacy amplification is implemented by a universal$_2$ hash functions, it can yield protocols for the above tasks with a relatively small amount of calculation.

Similar to the study [2], [30] for random privacy amplification based on the universal$_2$ condition, the previous paper[6] focused only on the mutual information with the eavesdropper. However, as the secrecy criterion, many papers in cryptography community [22], [3], [4], [5] adopt the half of the L_1 norm distance, which is also called the variation distance or Eve's distinguishability. Because this criterion is closely related to universally composable security [22], it is called the universal composability and is required to evaluate the leaked information based on the L_1 norm distance from cryptography community viewpoint.

In this paper, we adopt the L_1 norm distance as the secrecy criterion, i.e., the universal composability, and evaluate the secrecy for random privacy amplification. In the independent and identical distributed case, when the rate of generation random numbers is smaller than the entropy of the original information source, it is possible to generate the random variable whose L_1 norm distance to the uniform random number approaches zero asymptotically. However, in the realistic setting, we can manipulate only a finite size of random variables. So, the speed of this convergence is very important. In the community of information theory, in order to discuss the speed, we often focus on the the exponential rate of decrease. This rate is called the exponent, and is widely discussed among several topics in information theory, e.g., channel coding[20], source coding[13], [31], and mutual information criterion in wire-tap channel[17], [6]. However, the exponent has not been discussed in the community of cryptography as an important criterion. The purpose of this paper is establishing a systematic evaluating method for exponent for the L_1 norm distance in secure protocols.

In Section III, first, we focus on Bennett et al[2]'s evaluation for random privacy amplification, which employs the Rényi entropy of order 2. This evaluation was also obtained by Håstad et al [30] and is often called leftover hash lemma. Using a discussion similar to Renner [5], we derive an upper bound for the L_1 norm distance under the univeresal$_2$ condition for hash functions, which is the main theorem of this paper (Theorem 1).

Next, we apply this theorem to the i.i.d. setting with a given generating rate and a given source distribution. Then, we derive a lower bound of the exponent of the average of the L_1 norm distance between the generated random number and the uniform random number when a family of universal$_2$ hash functions is applied. Next, we introduce a stronger condition for hash functions, which is called strongly universal$_2$. We consider the n-independent and identical extension, and show that the exponential rate of decrease for this bound is tight under a stronger condition by using the type method[13], which was invented by Csiszár and Körner [13] and is one of standard methods in information theory. Since our bound realizes the optimal exponent, it is thought to be powerful even for the finite length setting. However, if our protocol generating the random number is allowed to depend on the original distribution, there is a possibility to improve the exponent while it is known that asymptotic generation cannot be improved[26]. In Section IV, we derive the optimal exponent in this setting by using the Cramer's Theorem [27] and the type method [13]. Comparing these two exponents, we can compare the performances between the protocol based on universal$_2$ hash functions and the protocol depending on the information

M. Hayashi is with Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai, 980-8579, Japan and Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542. (e-mail: hayashi@math.is.tohoku.ac.jp)

source.

In Section V, we consider the case when an eavesdropper has a random variable correlated to the random variable of the authorized user. In this case, when the authorized user applies universal$_2$ hash functions to his random variable, he obtain a secure random variable. When we apply Theorem 1 to the security by L_1 norm distance in this setting, we obtain a tighter evaluation (59) than existing evaluation than that directly obtained from the previous paper[6].

In Section VI, we focus on wire-tap channel model, whose capacity has been calculated by Wyner [12] and Csiszár and Körner [13]. Csiszár [14] showed the strong security, and many papers [6], [33], [34] treat this model with mutual information criterion. The previous paper [17] derived bounds for both exponential rates of decrease for the security criterion based on the L_1 norm distance as well as the mutual information between Alice and Eve. It obtained a bound for the exponential rate of decrease concerning the L_1 security criterion. In this paper, we apply (59) to wire-tap channel model, and obtain the evaluation of the exponent of the L_1 security criterion. In Section VII, it is shown that the evaluation obtained in this paper is better than that by the previous paper [17]. In a realistic setting, it is natural to restrict our codes to linear codes. In Section VIII, using (67), we provide a security analysis for a code constructed by the combination of an arbitrary linear code and the privacy amplification by universal$_2$ hash functions. This analysis yields the exponential rate of decrease for the L_1 security criterion. Overall, since (59) and (67) are derived from Theorem 1, all of the obtained results concerning the wire-tap channel model can be regarded as consequences of Theorem 1.

Further, in Section IX, we obtain the bound for the L_1 security criterion in one-way secret key generation. In Appendix A, we prove Theorem 2 mentioned in Subsection III-A. In Appendix B, we prove Lemma 6 given in Subsection IV.

Relation with the previous paper[6]

The main difference from the previous paper [6] is that the analysis on this paper is based on the universal composability while that on the previous paper [6] is based on the mutual information criterion. In the first step, this paper derives an evaluation (Theorem 1) of the equality of the uniform random number generation by universal$_2$ hash functions based on the L_1 norm criterion. Applying Theorem 1, we treat several security problems. Since this paper treats the same security problems as the previous paper with the different criterion, some of protocols used in this paper were used in the previous paper[6]. That is, the coding protocols used in Sections VI, VIII and IX are used in Sections III, V, and VI in [6], respectively While these protocols are described in [6], we describe the whole protocols in this paper for the readers' convenience.

For the uniform random number generation, this paper gives the tight exponential rate of decrease for the L_1 norm distance, while the previous paper[6] gives a lower bound of the exponential rate of decrease based on Shannon entropy. Concerning the secret key generation without communication, this paper gives a lower bound of the exponential rate of decrease based on the universal composability, while the previous paper[6] gives a lower bound of the exponential rate of decrease based on the mutual information criterion. Applying Pinsker inequality (5), we can derive a lower bound of the exponential rate of decrease based on the universal composability from the lower bound by [6]. As is shown in Lemma 8 in Subsection V-B, our lower bound is (strictly) better than combination of Pinsker inequality and the lower bound by [6] (except for special cases). Note that application of Pinsker inequality (5) or (6) yields the half of the lower bound of the exponent of the mutual information as a lower bound of the exponent of universal composability. Indeed, we give a numerical example at Fig. 2, in which, our bound is strictly better than that by [6].

Concerning wire-tap channel in a general framework, the code given in this paper is quite similar to that in the previous paper[6]. However, the evaluation method in this paper is different with that of the previous paper[6] because the analysis in this paper is based on the universal composability while that in the previous paper[6] is based on the mutual information. In this model, we can derive a lower bound for the exponential rate of decrease based on the universal composability by the combination of Pinsker inequality (5) and the result in [6]. As is shown in Section VII, our lower bound is better than this lower bound by [6]. Section VIII treats a more realistic setting by using linear codes. Even in this setting, as is explained in Remark 1, our lower bound is (strictly) better than the lower bound by [6] (except for special cases mentioned in Lemma 8). The same observation can be applied to secret key generation by public communication, which is discussed in Section IX.

II. Preliminaries

First, we briefly explain several notations and basic knowledge in information theory. In order to evaluate the difference two distributions P^X and $\tilde{P}^X$, we employ the following quantities: the L_1 distance (variational distance)

$$d_1(P^X, \tilde{P}^X) := \sum_x |P^X(x) - \tilde{P}^X(x)|, \tag{1}$$

the L_2 distance

$$d_2(P^X, \tilde{P}^X) := \sqrt{\sum_x (P^X(x) - \tilde{P}^X(x))^2}, \tag{2}$$

and the KL-divergence

$$D(P^X \| \tilde{P}^X) := \sum_x P^X(x)(\log P^X(x) - \log \tilde{P}^X(x)). \tag{3}$$

These definitions can be extended when the total measure is less than 1 i.e., $\sum_a P^A(a) \le 1$. In the following, we call such P^A a sub-distribution. This extension for sub-distributions is crucial for the later discussion.

When a joint distribution $P^{X,Y}$ is given, we have the

following equation

$$d_1(P^{X,Y}, \tilde{P}^X \times P^Y) = \sum_{x,y} |P^{X,Y}(x,y) - \tilde{P}^X(x)P^Y(y)|$$
$$= \sum_y P^Y(y) \sum_x |P^{X|Y}(x|y) - \tilde{P}^X(x)|$$
$$= \sum_y P^Y(y) d_1(P^{X|Y=y}, \tilde{P}^X). \quad (4)$$

When $P^X, \tilde{P}^X$ are normalized distributions, as a relation between the KL-divergence and the L_1 distance, the Pinsker inequality

$$d_1(P^X, \tilde{P}^X)^2 \le D(P^X \| \tilde{P}^X) \quad (5)$$

is known [19]. That is,

$$-\log d_1(P^X, \tilde{P}^X) \ge \frac{-1}{2} \log D(P^X \| \tilde{P}^X). \quad (6)$$

These relations will be helpful for the latter discussions.

III. Uniform random number generation

A. Protocol based on universal$_2$ hash function: Direct part

Firstly, we consider the uniform random number generation problem from a biased random number $a \in \mathcal{A}$, which obeys a probability distribution P^A when its cardinality $|\mathcal{A}|$ is finite. There are two types of protocols for this problem. One is a protocol specialized for the given distribution P^A. The other is a universal protocol that does not depends on the given distribution P^A. The aim of this section is evaluate the performance of the latter setting. In the latter setting, our protocol is given by a function f from $\mathcal{A}$ to $\mathcal{M} = \{1, \ldots, M\}$.

The quality of the random number obeying the sub-distribution P^A is evaluated by

$$d_1(P^A) := d_1(P^A, P^A(\mathcal{A})P^A_{\rm mix}), \quad (7)$$

where $P^A_{\rm mix}$ is the uniform distribution on $\mathcal{A}$. We also use the Rényi entropy order $1+s$:

$$H_{1+s}(A|P^A) := \frac{-1}{s} \log \sum_a P^A(a)^{1+s}.$$

The L_2 distance is written by using the Rényi entropy order 2 as follows.

$$d_2(P^A, P^A(\mathcal{A})P^A_{\rm mix})^2 = e^{-H_2(A|P^A)} - \frac{P^A(\mathcal{A})^2}{|\mathcal{A}|}. \quad (8)$$

Now, we focus on an ensemble of the functions $f_{\mathbf{X}}$ from $\mathcal{A}$ to $\mathcal{M} = \{1, \ldots, M\}$, where $\mathbf{X}$ denotes a random variable describing the stochastic behavior of the function $f_{\mathbf{X}}$. In this case, we adopt on the following quantity as a criterion of the secrecy:

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) = \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\rm mix})$$
$$= d_1(P^{B,\mathbf{X}}, P^A(\mathcal{A})P^B_{\rm mix} \times P^{\mathbf{X}}), \quad (9)$$

where B is the random variable $f_{\mathbf{X}}(A)$ and the final equation follows from (4). Hence, when the expectation $\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)})$ is sufficiently small, the random variable $f_{\mathbf{X}}(A)$ is almost independent of the side information $\mathbf{X}$. Then, the choice $f_{\mathbf{X}}$ can be communicated between Alice and Bob without revealing anything about $f(A)$.

An ensemble of the hash functions $f_{\mathbf{X}}$ is called universal$_2$ when it satisfies the following condition[1]:

Condition 1 (Universal$_2$): $\forall a_1 \neq \forall a_2 \in \mathcal{A}$, the collision probability that $f_{\mathbf{X}}(a_1) = f_{\mathbf{X}}(a_2)$ is at most $\frac{1}{M}$.

We sometimes require the following additional condition:

Condition 2: For any $\mathbf{X}$, the cardinality of $f_{\mathbf{X}}^{-1}\{i\}$ does not depend on i.

This condition will be used in Section IV.

Indeed, when the cardinality $|\mathcal{A}|$ is a power of a prime power q and M is another power of the same prime power q, as is shown in Appendix II of the previous paper [6], the ensemble $\{f_{\mathbf{X}}\}$ can be given by the concatenation of Toeplitz matrix and the identity $(\mathbf{X}, I)$[18] only with $\log_q |\mathcal{A}| - 1$ random variables taking values in the finite field $\mathbb{F}_q$. That is, the function can be obtained by the multiplication of the random matrix $(\mathbf{X}, I)$ taking values in $\mathbb{F}_q$. In this case, Condition 2 can be confirmed because the rank of $(\mathbf{X}, I)$ is constant.

Bennett et al[2] essentially showed the following lemma.

Lemma 1: A family of universal$_2$ hash functions $f_{\mathbf{X}}$ satisfies

$$\mathrm{E}_{\mathbf{X}} e^{-H_2(f_{\mathbf{X}}(A)|P^{f_{\mathbf{X}}(A)})} \le e^{-H_2(A|P^A)} + \frac{P^A(\mathcal{A})^2}{M}. \quad (10)$$

This was also shown by Håstad et al [30] and is often called leftover hash lemma.

Now, we follow the derivation of Theorem 5.5.1 of Renner [5] when one classical random variable is given. The Schwarz-inequality implies that

$$d_1(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\rm mix})$$
$$\le \sqrt{M} \sqrt{d_2(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\rm mix})}.$$

The Jensen inequality yields that

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\rm mix})$$
$$\le \sqrt{M} \sqrt{\mathrm{E}_{\mathbf{X}} d_2(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\rm mix})}.$$

Substituting (8) and (10) into the above inequality, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \le M^{\frac{1}{2}} e^{-\frac{H_2(A|P^A)}{2}}. \quad (11)$$

Using (11), we can show the following theorem as a generalization of (11).

Theorem 1: A family of universal$_2$ hash functions $f_{\mathbf{X}}$ satisfies

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \le 3M^{\frac{s}{1+s}} e^{-\frac{sH_{1+s}(A|P^A)}{1+s}} \text{ for } 0 \le \forall s \le 1. \quad (12)$$

Substituting $s = 1$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \le 3M^{\frac{1}{2}} e^{-\frac{H_2(A|P^A)}{2}}. \quad (13)$$

Since the difference between (11) and (13) is only the coefficient, Theorem 1 can be regarded as a kind of generalization of Bennett et al[2]'s result (10).

Proof: For any $R' > 0$, we choose subset $\Omega_{R'} := \{P^A(a) > e^{-R'}\}$, and define the sub-distribution $P^A_{R'}$ by

$$P^A_{R'}(a) := \begin{cases} 0 & \text{if } a \in \Omega_{R'} \\ P^A(a) & \text{otherwise.} \end{cases}$$

Since

$$d_1(P^A, P^A_{R'}) = P^A(\Omega_{R'})$$

and

$$\begin{aligned} & d_1(P^A_{R'}(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ =& d_1(0, (P^A(\mathcal{A}) - P^A_{R'}(\mathcal{A}))P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ =& (P^A(\mathcal{A}) - P^A_{R'}(\mathcal{A}))d_1(0, P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ =& P^A(\mathcal{A}) - P^A_{R'}(\mathcal{A}) = P^A(\Omega_{R'}), \end{aligned}$$

the idea of "smoothing" by Renner [5] yields that

$$\begin{aligned} & d_1(P^{f_{\mathbf{X}}(A)}) = d_1(P^{f_{\mathbf{X}}(A)}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ \le& d_1(P^{f_{\mathbf{X}}(A)}, P^{f_{\mathbf{X}}(A)}_{R'}) + d_1(P^{f_{\mathbf{X}}(A)}_{R'}, P^A_{R'}(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ & + d_1(P^A_{R'}(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}, P^A(\mathcal{A})P^{f_{\mathbf{X}}(A)}_{\mathrm{mix}}) \\ =& 2P^A(\Omega_{R'}) + d_1(P^{f_{\mathbf{X}}(A)}_{R'}). \end{aligned} \tag{14}$$

Taking the expectation concerning $\mathbf{X}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \le 2P^A(\Omega_{R'}) + \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}_{R'}). \tag{15}$$

The inequality (11) yields

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}_{R'}) \le M^{\frac{1}{2}} e^{-\frac{1}{2}H_2(A|P^A_{R'})}.$$

For $0 \le s \le 1$, we can evaluate $e^{-H_2(A|P^A_{R'})}$ and $P^A(\Omega_{R'})$ as

$$\begin{aligned} & e^{-H_2(A|P^A_{R'})} = \sum_{a \in \Omega^c_{R'}} P^A(a)^2 \le \sum_{a \in \Omega^c_{R'}} P^A(a)^{1+s} e^{-(1-s)R'} \\ \le& \sum_a P^A(a)^{1+s} e^{-(1-s)R'} = e^{-sH_{1+s}(A|P^A)-(1-s)R'} \end{aligned} \tag{16}$$

$$\begin{aligned} & P^A(\Omega_{R'}) = \sum_{a \in \Omega_{R'}} P^A(a) \le \sum_{a \in \Omega_{R'}} (P^A(a))^{1+s} e^{sR'} \\ \le& \sum_a (P^A(a))^{1+s} e^{sR'} = e^{-sH_{1+s}(A|P^A)+sR'}. \end{aligned} \tag{17}$$

Combining (15), (16), and (17), for $R := \log M$, we obtain

$$\begin{aligned} & \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \\ \le& 2e^{-sH_{1+s}(A|P^A)+sR'} + e^{R+\frac{1}{2}(-sH_{1+s}(A|P^A)-(1-s)R')} \\ =& 3e^{-\frac{sH_{1+s}(A|P^A)+sR}{1+s}}, \end{aligned}$$

where we substitute $\frac{R+sH_{1+s}(A|P^A)}{1+s}$ into R'. ∎

Next, we consider the case when our distribution P^{A_n} is given by the n-fold independent and identical distribution of P^A, i.e, $(P^A)^n$. When the random number generation rate $\lim_{n\to\infty} \frac{1}{n} \log M_n$ is R, we focus on the *exponential rate of decrease* of $\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)})$, and consider the supremum.

When an ensemble $\{f_{\mathbf{X},n}\}$ of hash functions is a family of universal$_2$ hash functions from $\mathcal{A}^n$ to $\{1, \dots M_n\}$, Theorem 1 yields that

$$\begin{aligned} & \liminf_{n\to\infty} \frac{-1}{n} \log \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \\ \ge& \frac{sH_{1+s}(A|P^A) - sR}{1+s} \end{aligned}$$

for $s \in [0, 1]$. Taking the maximum concerning $s \in [0, 1]$, we obtain

$$\begin{aligned} & \liminf_{n\to\infty} \frac{-1}{n} \log \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \\ \ge& \max_{0\le s\le 1} \frac{sH_{1+s}(A|P^A) - sR}{1+s}. \end{aligned} \tag{18}$$

On the other hand, when we apply the Pinsker inequality[19] to the upper bound for the mutual information obtained by the previous paper [6], we obtain another bound $\max_{0\le s\le 1} \frac{sH_{1+s}(A|P^A)-sR}{2}$, which is smaller than (18).

B. Protocol based on universal$_2$ hash function: Converse part

In oder to show the tightness of the exponential rate of decrease (18) under the universal$_2$ condition, we consider the following ensemble.

Condition 3 (Strongly universal$_2$): For any $a \in \mathcal{A}$, $\Pr\{f_{\mathbf{X}}(a) = m\} = \frac{1}{M}$. The random variable $f_{\mathbf{X}}(a)$ is independent of $\{f_{\mathbf{X}}(a')\}_{a' \ne a \in \mathcal{A}}$.

Theorem 2: Under the strongly universal$_2$ ensemble, and any subset $\Omega \subset \mathcal{A}$ with $|\Omega| < M$ satisfies

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \ge (1 - \frac{|\Omega|}{M})^2 P^A(\Omega). \tag{19}$$

Its proof is given in Appendix A.

In order to derive the inequality opposite to (18) from Theorem 2, we employ the type method[19]. In the type method, when an n-trial data $\vec{a}_n := (a_1, \dots, a_n) \in \mathcal{A}^n$ is given, we focus on the distribution $p(a) := \frac{\#\{i|a_i=a\}}{n}$, which is called the empirical distribution for the data $\vec{a}_n$. In the type method, an empirical distribution is called a type. In the following, we denote the set of empirical distributions on $\mathcal{A}$ with n trials by $\mathcal{T}_n$. The cardinality $|\mathcal{T}_n|$ is bounded by $(n+1)^{|\mathcal{A}|-1}$ [19], which increases polynomially concerning the number n. That is,

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{T}_n| = 0. \tag{20}$$

This property is the key idea in the type method. When $T_n(Q)$ represents the set of n-trial data whose empirical distribution is Q, the cardinality of $T_n(Q)$ can be evaluated as [19]:

$$\lceil \frac{e^{nH(Q)}}{|\mathcal{T}_n|} \rceil \le |T_n(Q)| \le \lfloor e^{nH(Q)} \rfloor, \tag{21}$$

where $\lceil x \rceil$ is the minimum m satisfying $m \ge x$, and $\lfloor x \rfloor$ is the maximum m satisfying $m \le x$. Since any element $\vec{a} \in T_n(Q)$ satisfies

$$P^{A_n}(\vec{a}) = e^{-n(D(Q\|P^A)+H(Q))}, \tag{22}$$

we obtain an important formula

$$\frac{1}{\mathcal{T}_n} e^{-nD(Q\|P^A)} \le P^{A_n}(T_n(Q)) \le e^{-nD(Q\|P^A)}. \tag{23}$$

Using the above knowledge, we can show the following proposition:

Proposition 1: When $M_n = \lfloor e^{nR} \rfloor$, any sequence of strongly universal$_2$ ensemble $\{f_{\mathbf{X},n}\}$ from $\mathcal{A}^n$ to $\{1, \ldots M_n\}$ satisfies the equation

$$\limsup_{n\to\infty} \frac{-1}{n} \log \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \le \min_{Q:H(Q)\le R} D(Q\|P^A), \tag{24}$$

where $D(Q\|P^A)$ is the Kullback-Leibler divergence $\sum_{a\in\mathcal{A}} Q(a)(\log Q(a) - \log P^A(a))$.

Proof: Choose an arbitrary empirical distribution $Q \in \mathcal{T}_n$ satisfying that $H(Q) \le R$. Then, due to (21), the cardinality $|T_n(Q)|$ is less than $\lfloor e^{nR} \rfloor$. We choose the subset $\Omega_{n,Q}$ with the cardinality $\lceil \frac{1}{2} e^{nR} \rceil$ so that it at least contains $\lceil \frac{|T_n(Q)|}{2} \rceil$ elements of $T_n(Q)$. Using (21) and (22), we obtain

$$\begin{aligned} &P^{A_n}(\Omega_{n,Q}) \ge \frac{|T_n(Q)|}{2} e^{-n(D(Q\|P^A)+H(Q))} \\ \ge &\frac{e^{nH(Q)}}{2|\mathcal{T}_n|} e^{-n(D(Q\|P^A)+H(Q))}. \end{aligned}$$

Using Theorem 2 with $\Omega_{n,Q}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \ge (1 - \frac{\lceil \frac{1}{2} e^{nR} \rceil}{\lfloor e^{nR} \rfloor})^2 \frac{1}{2|\mathcal{T}_n|} e^{-nD(Q\|P^A)}.$$

Since Q is an arbitrary empirical distribution $Q \in \mathcal{T}_n$ satisfying that $H(Q) \le R$,

$$\begin{aligned} &\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \\ \ge &(1 - \frac{\lceil \frac{1}{2} e^{nR} \rceil}{\lfloor e^{nR} \rfloor})^2 \frac{1}{2|\mathcal{T}_n|} \max_{Q\in\mathcal{T}_n:H(Q)\le R} e^{-nD(Q\|P^A)}. \end{aligned}$$

That is,

$$\begin{aligned} &\frac{-1}{n} \log \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \\ \le &\min_{Q\in\mathcal{T}_n:H(Q)\le R} D(Q\|P^A) + \frac{1}{n} \log 2|\mathcal{T}_n| \\ &- \frac{2}{n} \log(1 - \frac{\lceil \frac{1}{2} e^{nR} \rceil}{\lfloor e^{nR} \rfloor}). \end{aligned}$$

Due to the continuity of $Q \mapsto H(Q), (Q\|P^A)$ and (20), the limit $n \to \infty$ yields (24). ∎

When $R \le H(A|P^A)$, the equation

$$\max_{0\le s} \frac{s(H_{1+s}(A|P^A) - R)}{1+s} = \min_{Q:H(Q)\le R} D(Q\|P^A) \tag{25}$$

is known as the strong converse exponent in the fixed source coding [19], [13], [31],[24, (A21)]. The maximum $\max_{0\le s} \frac{s(H_{1+s}(A|P^A)-R)}{1+s}$ is realized at $s = s_0$ when $R = R_{s_0} := (1+s_0)\frac{d}{ds}(sH_{1+s}(A|P^A))|_{s=s_0} - s_0 H_{1+s_0}(A|P^A)$. Since $\frac{d}{ds} R_s = (1+s)\frac{d^2}{ds^2}(sH_{1+s}(A|P^A)) \le 0$, R_s is monotone decreasing concerning s.

Thus, when $H(A|P^A) \ge R \ge R_1$ (R_1 is called the critical rate.),

$$\max_{0\le s} \frac{s(H_{1+s}(A|P^A) - R)}{1+s} = \max_{0\le s\le 1} \frac{s(H_{1+s}(A|P^A) - R)}{1+s}. \tag{26}$$

Hence, in this case, due to (18), (24), (25), and (26), we obtain

$$\begin{aligned} &\lim_{n\to\infty} \frac{-1}{n} \log \mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X},n}(A_n)}) \\ = &\max_{0\le s\le 1} \frac{s(H_{1+s}(A|P^A) - R)}{1+s} = \min_{Q:H(Q)\le R} D(Q\|P^A). \end{aligned} \tag{27}$$

However, when $R < R_1$,

$$\begin{aligned} &\max_{0\le s\le 1} \frac{s(H_{1+s}(A|P^A) - R)}{1+s} = \frac{H_2(A|P^A) - R}{2} \\ < &\max_{0\le s} \frac{s(H_{1+s}(A|P^A) - R)}{1+s}. \end{aligned}$$

The lower bound in (18) does not coincide with the upper bound in (24).

C. Comparison with evaluation by Holenstein-Renner [29]

In the above derivation, the key point is evaluating the probability $P^A(\Omega_{R'})$, which equals the probability $(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\}$ in the n-i.i.d. setting. In the community of cryptography, the n-i.i.d. setting is not regarded as an important setting because they are more interested in the single-shot setting. In such a setting, they sometimes use Holenstein-Renner [29] evaluation for $P^X(\Omega_{R'})$. They proved the following theorem.

Theorem 3: When $\le H(A) - R' \le \log|\mathcal{A}|$,

$$(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} \le 2^{-\frac{n(H(A)-R')^2}{2(\log(|\mathcal{A}|+3))^2}}. \tag{28}$$

Further, When $|\mathcal{A}| \ge 3$ and $0 \le H(A) - R' \le \frac{\log(|\mathcal{A}|-1)}{12}$,

$$(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} > \frac{1}{110} 2^{-\frac{12n(H(A)-R')^2}{(\log(|\mathcal{A}|-1))^2}}.$$

When $|\mathcal{A}| = 2$, the inequality yields the following evaluation. When $0 \le H(A) - R' \le \frac{\log 3}{24}$,

$$(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} > \frac{1}{110} 2^{-\frac{24n(H(A)-R')^2}{(\log 3)^2}}$$

for even n.

Our evaluation (17) of $(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\}$ contains the parameter $0 \le s \le 1$. Since this parameter is arbitrary, it is natural to compare the upper bound $\min_{0\le s\le 1} e^{-n(sH_{1+s}(X|P^X)-sR')}$ given by (17) with that by Theorem 3. That is, using (17), we obtain the exponential evaluation

$$\begin{aligned} &\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} \\ \ge &\max_{0\le s} sH_{1+s}(A|P^A) - sR', \end{aligned}$$

while Theorem 3 yields that

$$\begin{aligned} &\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} \\ \ge &\frac{(H(A)-R')^2}{2(\log(|\mathcal{A}|+3))^2} \log 2. \end{aligned}$$

In this case, the upper bound is $\frac{12\log 2(H(A)-R')^2}{(\log(|\mathcal{A}|-1))^2}$ for $|\mathcal{A}| \ge 3$ and $\frac{24\log 2(H(A)-R')^2}{(\log 3)^2}$ for $|\mathcal{A}| = 2$.

In fact, the probability $P^A(\Omega_{R'})$ is the key quantity in the method of information spectrum, which is a unified method in information theory[32]. When the method of information spectrum is applied to the i.i.d. source, the probability $P^A(\Omega_{R'})$ is evaluated by applying Cramér Theorem (See [27]) to the random variable $\log P^A(a)$. Then, we obtain

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} \\ =& \max_{0\le s} sH_{1+s}(A|P^A) - sR' \end{aligned} \tag{29}$$

for $R \le H(A)$. Since $s \mapsto sH_{1+s}(X|P^X)$ is concave, when $H(A) \ge R \ge H_2'(A|P^A)$, the maximization (29) can be attained with $s \in [0,1]$, i.e.,

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\} \\ =& \max_{0\le s\le 1} sH_{1+s}(A|P^A) - sR', \end{aligned}$$

which implies that our evaluation (17) gives the tight bound for exponential rate of decrease for the probability $(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\}$. In fact, the difference among these bounds is numerically given in Fig. 1. Therefore, we can conclude that our evaluation (17) is much better than that by Holenstein-Renner [29]. That is, the combination of Lemma 1 and (17) is essential for deriving the tight exponential bound.

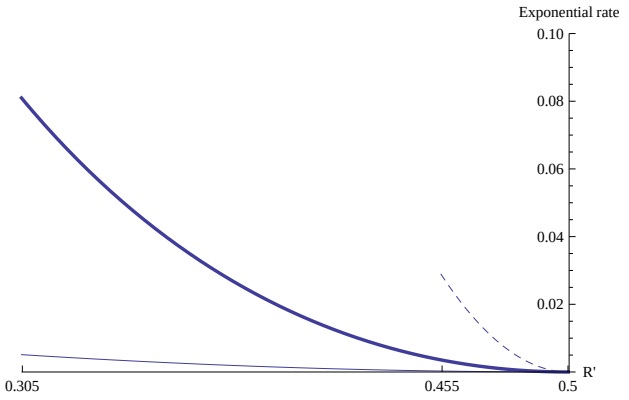

Fig. 1. Evaluation of $\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) > e^{-nR'}\}$. Thick line: $\max_{0\le s\le 1} sH_{1+s}(A|P^A) - sR'$ (The present paper), Normal line: $\frac{(H(A)-R')^2}{2(\log(|\mathcal{A}|+3))^2} \log 2$ (Lower bound by [29]), Dashed line: $\frac{24 \log 2 (H(A)-R')^2}{(\log 3)^2}$ (Upper bound by [29]) $p = 0.200$, $h(p) = H(A) = 0.500$, $\frac{d(sH_{1+s}(A))}{ds}|_{s=1} = 0.305$, $H(A) - \frac{\log 3}{24} = 0.455$.

IV. Specialized protocol for uniform random number generation

A. Main result of this section

Next, we consider a function f from $\mathcal{A}$ to $\{1,\ldots,M\}$ specialized to a given probability distribution P^A. This problem is called intrinsic randomness, which was studied with general source by Vembu and Verdú [26]. The previous paper [25] discussed the relation between the second order asymptotic rate and the central limit theorem. In the following, for the comparison with the exponential rate of decrease for (25), we prove the following theorem, which gives the optimal exponential rate of decrease for a given rate generating uniform random number.

Theorem 4: When $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} \le R$, we obtain

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log \min_{f_n \in \mathcal{F}_n(R)} d_1(P^{f_n(A_n)}) \\ =& \max_{0\le s\le 1} s(H_{1+s}(A|P^A) - R), \end{aligned} \tag{30}$$

where $\mathcal{F}_n(R)$ is the set of functions f_n from $\mathcal{A}^n$ to $\{1,\ldots,\lfloor e^{nR} \rfloor\}$.

Combining (27) and Theorem 4, we can compare the performances between a random universal protocol and the best specialized protocol. So, our exponential rate of decrease for the protocol based on universal$_2$ hash functions is slightly smaller than the optimal exponential rate of decrease for specialized protocols.

In order to prove Theorem 4, we will show the following two inequalities:

$$\begin{aligned}&\limsup_{n\to\infty} \frac{-1}{n} \log \min_{f_n \in \mathcal{F}_n(R)} d_1(P^{f_n(A_n)}) \\ \le& \max_{0\le s\le 1} s(H_{1+s}(A|P^A) - R) \end{aligned} \tag{31}$$

$$\begin{aligned}&\liminf_{n\to\infty} \frac{-1}{n} \log \min_{f_n \in \mathcal{F}_n(R)} d_1(P^{f_n(A_n)}) \\ \ge& \max_{0\le s\le 1} s(H_{1+s}(A|P^A) - R). \end{aligned} \tag{32}$$

Inequality (31) is called the converse part and Inequality (32) is called the direct part in information theory community. In order to show respective inequalities, we prepare respective lemmas (Lemmas 2 and 4) with non-asymptotic setting in Subsection IV-B. In Subsection IV-C, using Lemma 4 and the concavity property, we show the converse part (31). Also, using Lemma 2, we show the direct part (31). In the latter derivation, we employ the method of type, which is one of standard method in information theory[19].

B. Non-asymptotic evaluation

In order to treat the non-asymptotic case, we introduce the notation:

$$[x]_+ := \begin{cases} x & \text{if } x \ge 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Then, the L_1 norm for two normalized distributions P and Q can be simplified to

$$\sum_a |P(a) - Q(a)| = 2 \sum_a [P(a) - Q(a)]_+, \tag{33}$$

which is a useful formula for the following discussion.

Hence, we obtain the following lemma, which is useful for our proof of the direct part (32).

Lemma 2: Any probability distribution P^A and any function f from $\mathcal{A}$ to $\{1,\ldots,M\}$ satisfy that

$$d_1(P^{f(A)}) \ge P^A\{a \in \mathcal{A} | P^A(a) \ge \frac{2}{M}\}. \tag{34}$$

Proof:

Any positive numbers $\alpha_1, \ldots, \alpha_k$ satisfies

$$[\sum_{i=1}^{k} \alpha_i - \frac{1}{M}]_+ \ge \sum_{i=1}^{k} [\alpha_i - \frac{1}{M}]_+. \tag{35}$$

When $P^A(a) \ge \frac{2}{M}$, $P^A(a) - \frac{1}{M} \ge \frac{1}{M}$, which implies that

$$\begin{aligned} 2[P^A(a) - \frac{1}{M}]_+ &= 2(P^A(a) - \frac{1}{M}) \\ &\ge P^A(a) - \frac{1}{M} + \frac{1}{M} = P^A(a). \end{aligned} \tag{36}$$

Thus, we obtain

$$\begin{aligned} &\sum_b |P^A(f^{-1}(b)) - \frac{1}{M}| = 2 \sum_b [P^A(f^{-1}(b)) - \frac{1}{M}]_+ \\ \ge& 2 \sum_{a \in \mathcal{A}} [P^A(a) - \frac{1}{M}]_+ \qquad (37) \\ \ge& 2 \sum_{a \in \mathcal{A}: P^A(a) \ge \frac{2}{M}} [P^A(a) - \frac{1}{M}]_+ \\ \ge& \sum_{a \in \mathcal{A}: P^A(a) \ge \frac{2}{M}} P^A(a), \qquad (38) \end{aligned}$$

where (37) and (38) follows from (35) and (36). Therefore, we obtain (34). ∎

In order to show the converse part, we prepare the following lemma.

Lemma 3: Assume that for two integers $M \ge N$, two positive number sequences $\alpha_1, \ldots, \alpha_N$ and $\beta_1, \ldots, \beta_M$ satisfy that $\sum_{i=1}^N \alpha_i \ge \sum_{i=1}^M \beta_i$. Then, there exists a map f from $\{1, \ldots, M\}$ to $\{1, \ldots, N\}$ such that

$$\sum_{i=1}^{N} [\sum_{j \in f^{-1}(i)} \beta_j - \alpha_i]_+ \le N \max_j \beta_j. \tag{39}$$

Proof: First, we define $f(1) := 1$. For $j > 1$, we define $f(j)$ inductively. When $\sum_{j' \in f^{-1}(f(j-1))} \beta_{j'} < \alpha_{f(j-1)}$. we define $f(j) := f(j-1)$. Otherwise, we define $f(j) := f(j-1) + 1$. Then, the function satisfies the condition (39). ∎

Now, we consider the case when our distribution P^{A_n} is given by the n-fold independent and identical distribution of P^A, i.e, $(P^A)^n$. Using Lemma 3, we have the following lemma, which is useful for our proof of the converse part (31).

Lemma 4: For any probability distribution P^A, there exists a function f_n from $\mathcal{A}^n$ to $\{1, \ldots, M_n\}$ such that

$$\begin{aligned} &d_1(P^{f_n(A_n)}) \\ \le& 2(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\} \\ &+ 2 \sum_{Q \in \mathcal{T}_n^1[M_n]} M_n e^{-n(D(Q\|P^A)+H(Q))} \cdot (P^A)^n(T_n(Q)) \\ &+ 2|\mathcal{T}_n| \max_{Q \in \mathcal{T}_n^2[M_n]} e^{-n(D(Q\|P^A)+H(Q))} \end{aligned} \tag{40}$$

where

$$\begin{aligned} \mathcal{T}_n^1[M_n] &:= \{Q \in \mathcal{T}_n | D(Q\|P^A) + H(Q) \ge \frac{1}{n} \log M_n\} \\ \mathcal{T}_n^2[M_n] &:= \{Q \in \mathcal{T}_n | (P^A)^n(T_n(Q)) < \frac{1}{M_n}\}. \end{aligned}$$

Proof: In the first step, we define the function f_n. In the second step, we show that the function satisfies (40).

we divide $\mathcal{T}_n$ into three parts:

$$\begin{aligned} \tilde{\mathcal{T}}_n^0[M_n] &:= \{Q \in \mathcal{T}_n | e^{n(D(Q\|P^A)+H(Q))} \le M_n\} \\ \tilde{\mathcal{T}}_n^1[M_n] &:= \{Q \in (\tilde{\mathcal{T}}_n^0[M_n])^c \cap \mathcal{T}_n | (P^A)^n(T_n(Q)) \ge \frac{1}{M_n}\} \\ \tilde{\mathcal{T}}_n^2[M_n] &:= \{Q \in (\tilde{\mathcal{T}}_n^0[M_n])^c \cap \mathcal{T}_n | (P^A)^n(T_n(Q)) < \frac{1}{M_n}\}, \end{aligned}$$

where $(\tilde{\mathcal{T}}_n^0[M_n])^c$ is the compliment of $\tilde{\mathcal{T}}_n^0[M_n]$. These three parts have the following relation with the above two parts:

$$\tilde{\mathcal{T}}_n^1[M_n] \subset \mathcal{T}_n^1[M_n], \quad \tilde{\mathcal{T}}_n^2[M_n] \subset \mathcal{T}_n^2[M_n]$$

By using the integer $n_Q := \lfloor \frac{(P^A)^n(T_n(Q))}{1/M_n} \rfloor = \lfloor M_n (P^A)^n(T_n(Q)) \rfloor$, the conditions for $\tilde{\mathcal{T}}_n^1[M_n]$ and $\tilde{\mathcal{T}}_n^2[M_n]$ are written as $n_Q \ge 1$ and $n_Q < 1$, respectively. Note that, since n_Q is a non-negative integer, $n_Q < 1$ is equivalent with $n_Q = 0$.

Due to (22), the condition that $e^{n(D(Q\|P^A)+H(Q))} \le M_n$ is equivalent with the condition that $P^{A_n}(a) \ge \frac{1}{M_n}$ for $a \in T_n(Q)$. Hence,

$$(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\} = \sum_{Q \in \mathcal{T}_n^0} (P^A)^n(T_n(Q)). \tag{41}$$

So,

$$\begin{aligned} &(P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\} + \sum_{Q \in \tilde{\mathcal{T}}_n^1[M_n]} \frac{n_Q}{M_n} \\ \le& \sum_{Q \in \tilde{\mathcal{T}}_n^0[M_n]} (P^A)^n(T_n(Q)) + \sum_{Q \in \tilde{\mathcal{T}}_n^1[M_n]} (P^A)^n(T_n(Q)) \le 1. \end{aligned}$$

Since

$$\begin{aligned} &\frac{1}{M_n} \sum_{Q \in \tilde{\mathcal{T}}_n^0[M_n]} |T_n(Q)| = \frac{1}{M_n} |\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\}| \\ \le& (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\}, \end{aligned}$$

we have

$$\sum_{Q \in \tilde{\mathcal{T}}_n^0[M_n]} |T_n(Q)| + \sum_{Q \in \tilde{\mathcal{T}}_n^1[M_n]} n_Q \le M_n.$$

Therefore, we can choose f_n' on $\Omega' := \cup_{Q \in \tilde{\mathcal{T}}_n^0[M_n] \cup \tilde{\mathcal{T}}_n^1[M_n]} T_n(Q)$ satisfying the following conditions.

1) For $Q, Q' \in \tilde{\mathcal{T}}_n^0[M_n] \cup \tilde{\mathcal{T}}_n^1[M_n]$, $f_n'(T_n(Q)) \cap f_n'(T(Q')) = \emptyset$.
2) $f_n'|_{T_n(Q)}$ is injective for $Q \in \tilde{\mathcal{T}}_n^0[M_n]$.
3) $|f_n'(T_n(Q))| = n_Q$ for $Q \in \tilde{\mathcal{T}}_n^1[M_n]$. Further, we choose f_n' satisfying the additional condition.
4) Any type $Q \in \tilde{\mathcal{T}}_n^1[M_n]$ satisfies that $|{f_n'}^{-1}(b)| \le \frac{|T_n(Q)|}{n_Q}$ for $b \in f_n'(T_n(Q))$.

Then, for $Q \in \tilde{\mathcal{T}}_n^1[M_n]$, we obtain

$$P^{f_n'(A_n)}(b) \le \frac{1}{M_n} + e^{-n(D(Q\|P^A)+H(Q))}, \ \forall b \in f_n'(T_n(Q)). \tag{42}$$

From the construction,

$$\sum_{b\in f_n'(\Omega')} P^{f_n'(A_n)}(b) \ge \frac{1}{M_n}|f_n'(\Omega')|.$$

That is,

$$\sum_{a\in(\Omega')^c} P^{A_n}(a) \le \frac{1}{M_n}|(f_n'(\Omega'))^c|. \tag{43}$$

Next, we define f_n on the whole set by modifying f_n' as follows.

5) f_n is the same as f_n' on Ω'.
6) Due to (43), we can apply Lemma 3 to the case when $\{1,\ldots,N\} = (f_n'(\Omega'))^c$, $\{1,\ldots,M\} = (\Omega')^c$, $\alpha_b = \frac{1}{M_n}$ for $b \in (f_n'(\Omega'))^c$ and $\beta_a = P^{A_n}(a)$ for $a \in (\Omega')^c$. Following Lemma 3, we define the map $f_n|_{(\Omega')^c}$ from $(\Omega')^c$ to $(f_n'(\Omega'))^c$.

Our remaining task is to evaluate the value $\sum_b [P^{f_n(A_n)}(b) - \frac{1}{M_n}]_+$. Now, we define

$$C(Q) := \sum_{b\in f_n(T_n(Q))} [P^{f_n(A_n)}(b) - \frac{1}{M_n}]_+.$$

Then, (41) implies that

$$\sum_{Q\in\tilde{\mathcal{T}}_n^0[M_n]} C(Q) \le (P^A)^n\{a\in\mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\}. \tag{44}$$

For $Q \in \tilde{\mathcal{T}}_n^1[M_n]$, (42) implies

$$\begin{aligned} C(Q) \le & n_Q e^{-nD(Q\|P^A)-nH(Q)} \\ \le & M_n e^{-nD(Q\|P^A)-nH(Q)} \cdot (P^A)^n(T_n(Q)). \end{aligned} \tag{45}$$

Thus, (44) and (45) imply

$$\begin{aligned} & \sum_{b\in f_n'(\Omega')} [P^{f_n(A_n)}(b) - \frac{1}{M_n}]_+ \\ \le & (P^A)^n\{a\in\mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{M_n}\} \\ & + \sum_{Q\in\mathcal{T}_n^1[M_n]} M_n e^{-nD(Q\|P^A)-nH(Q)} \cdot (P^A)^n(T_n(Q)). \end{aligned} \tag{46}$$

Recall the condition 6). Lemma 3 guarantees that

$$\begin{aligned} & \sum_{b\in(f_n'(\Omega'))^c} [P^{f_n(A_n)}(b) - \frac{1}{M_n}]_+ \\ \le & |(f_n'(\Omega'))^c| \max_{Q\in\tilde{\mathcal{T}}_n^2[M_n]} e^{-n(D(Q\|P^A)+H(Q))} \\ \le & |\mathcal{T}_n| \max_{Q\in\mathcal{T}_n^2[M_n]} e^{-n(D(Q\|P^A)+H(Q))}. \end{aligned} \tag{47}$$

Combining (46) and (47), we obtain (40). ∎

C. Asymptotic evaluation

Next, we proceed to the asymptotic evaluation. First, using Cramer's theorem[27], we obtain

$$\begin{aligned} & \max_{0\le s} sH_{1+s}(A|P^A) - sR \\ = & \lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a\in\mathcal{A}^n | (P^A)^n(a) \ge \frac{1}{e^{nR}}\} \end{aligned} \tag{48}$$

Hence, Equality (48) and Lemma 2 imply

$$\begin{aligned} & \limsup_{n\to\infty} \frac{-1}{n} \log \min_{f_n\in\mathcal{F}_n(R)} d_1(P^{f_n(A_n)}) \\ \le & \max_{0\le s} s(H_{1+s}(A|P^A) - R). \end{aligned} \tag{49}$$

Since $s \mapsto sH_{1+s}(A|P^A)$ is concave, when $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} \le R$, the maximum $\max_{0\le s} s(H_{1+s}(A|P^A) - R)$ is realized at $s \in [0,1]$, i.e., $\max_{0\le s\le 1} s(H_{1+s}(A|P^A) - R) = \max_{0\le s} s(H_{1+s}(A|P^A) - R)$. Therefore, we obtain the converse part (31).

In order to show the direct part (32), we will show the following lemma by employing Lemma 2.

Lemma 5:

$$\begin{aligned} & \liminf_{n\to\infty} \frac{-1}{n} \log \min_{f_n\in\mathcal{F}_n(R)} d_1(P^{f_n(A_n)}) \\ \ge & \max_{0\le s\le 1} s(H_{1+s}(A|P^A) - R). \end{aligned} \tag{50}$$

In order to show Lemma 5, we prepare the following lemma, whose proof is given in Appendix B.

Lemma 6: When $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} \le R$,

$$\begin{aligned} & \min_{Q:H(Q)+D(Q\|P)\ge R} H(Q) + 2D(Q\|P) - R \\ = & \max_{0\le s} sH_{1+s}(A|P) - sR \\ = & \max_{0\le s\le 1} sH_{1+s}(A|P) - sR. \end{aligned} \tag{51}$$

When $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} > R$,

$$\begin{aligned} & \min_{Q:H(Q)+D(Q\|P)\ge R} H(Q) + 2D(Q\|P) - R \\ = & H_2(A|P) - R \end{aligned} \tag{52}$$

$$= \max_{0\le s\le 1} sH_{1+s}(A|P) - sR. \tag{53}$$

Proof of Lemma 5: Due to (20), (21), and the continuity of $Q \mapsto H(Q)$ and $D(Q\|P^A)$, we obtain

$$\begin{aligned} & \lim_{n\to\infty} \frac{-1}{n} \log 2|\mathcal{T}_n| \max_{Q\in\mathcal{T}_n^2[\lfloor e^{nR}\rfloor]} e^{-n(D(Q\|P^A)+H(Q))} \\ = & \lim_{n\to\infty} \min_{Q\in\mathcal{T}_n^2[\lfloor e^{nR}\rfloor]} D(Q\|P^A) + H(Q) \\ = & \min_{Q:D(Q\|P^A)\ge R} D(Q\|P^A) + H(Q) \\ \ge & \min_{Q:D(Q\|P^A)\ge R} H(Q) + 2D(Q\|P^A) - R \\ \ge & \min_{Q:H(Q)+D(Q\|P^A)\ge R} H(Q) + 2D(Q\|P^A) - R. \end{aligned} \tag{54}$$

From (23),

$$K_n := \sum_{Q\in\mathcal{T}_n^1[\lfloor e^{nR}\rfloor]} \lfloor e^{nR}\rfloor (P^A)^n(T_n(Q)) e^{-n(D(Q\|P^A)+H(Q))}$$

satisfies that

$$\begin{aligned} & \max_{Q\in\mathcal{T}_n^1[\lfloor e^{nR}\rfloor]} \frac{1}{\mathcal{T}_n} e^{-n(2D(Q\|P^A)+H(Q)-R)} \\ \le & K_n \le \mathcal{T}_n \max_{Q\in\mathcal{T}_n^1[\lfloor e^{nR}\rfloor]} e^{-n(2D(Q\|P^A)+H(Q)-R)}. \end{aligned}$$

Due to (20) and the continuity of $Q \mapsto H(Q)$ and $D(Q\|P^A)$,

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log K_n \\ =& \min_{Q:H(Q)+D(Q\|P^A)\ge R} H(Q)+2D(Q\|P^A)-R. \end{aligned} \quad (55)$$

As is shown in Lemma 6, RHSs of (54) and (55) equal $\max_{0\le s\le 1} sH_{1+s}(A|P^A) - sR$. Since $\max_{0\le s} sH_{1+s}(A|P^A) - sR \ge \max_{0\le s\le 1} sH_{1+s}(A|P^A) - sR$, (48) implies that

$$\begin{aligned}&\lim_{n\to\infty} \frac{-1}{n} \log (P^A)^n\{a \in \mathcal{A}^n | (P^A)^n(a) \ge \frac{2}{e^{nR}}\} \\ \ge& \max_{0\le s\le 1} s(H_{1+s}(A|P^A)-R). \end{aligned} \quad (56)$$

Thus, applying (54), (55), and (56) to the RHS of (40), and using Lemma 6, we can choose a sequence $\{f_n\}$ such that

$$\begin{aligned}&\liminf_{n\to\infty} \frac{-1}{n} \log \min_{f_n} d_1(P^{f_n(A_n)}) \\ \ge& \max_{0\le s\le 1} s(H_{1+s}(A|P^A)-R), \end{aligned} \quad (57)$$

which implies (50). ∎

V. Secret key generation without communication

A. Application of Theorem 1

Next, we consider the secure key generation problem from a common random number $A \in \mathcal{A}$ which has been partially eavesdropped on by Eve. For this problem, it is assumed that Alice and Bob share a common random number $A \in \mathcal{A}$, and Eve has another random number $E \in \mathcal{E}$, which is correlated to the random number A. The task is to extract a common random number $f(A)$ from the random number $A \in \mathcal{A}$, which is almost independent of Eve's random number $E \in \mathcal{E}$. Here, Alice and Bob are only allowed to apply the same function f to the common random number $A \in \mathcal{A}$.

Then, when the initial random variables A and E obey the distribution $P^{A,E}$, Eve's distinguishability can be represented by the following value:

$$d_1(P^{f(A),E}|E) := d_1(P^{f(A),E}, P_{\mathrm{mix}}^{f(A)} \times P^E),$$

where $P_{\mathrm{mix}}^{f(A)} \times P^E$ is the product distribution of both marginal distributions $P_{\mathrm{mix}}^{f(A)}$ and P^E, and $P_{\mathrm{mix}}^{f(A)}$ is the uniform distribution on $\{1,\ldots,M\}$. While the half of this value directly gives the probaility that Eve can distinguish the Alice's information, we call $d_1(P^{f(A),E}|E)$ Eve's distinguishability in the following. This criterion was proposed by [22] and was used by [5]. Since the half of this quantity $d_1(P^{f(A),E}|E)$ is closely related to the universally composable security, we adopt it as the secrecy criterion in this paper. As another criterion, we sometimes treat

$$d_1'(P^{f(A),E}|E) := d_1(P^{f(A),E}, P^{f(A)} \times P^E).$$

Since $d_1(P^{f(A)} \times P^E, P_{\mathrm{mix}}^M \times P^E) = d_1(P^{f(A)}, P_{\mathrm{mix}}^M) \le d_1(P^{f(A),E}, P_{\mathrm{mix}}^M \times P^E)$, we have

$$d_1'(P^{f(A),E}|E) \le 2 d_1(P^{f(A),E}|E).$$

Further, when $P^{f(A)}$ is the uniform distribution, the above both criteria coincide with each other.

Next, we consider an ensemble of universal$_2$ hash functions $\{f_{\mathbf{X}}\}$. Similar to (9), the equation

$$\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A),E}|E) = d_1(P^{B,E,\mathbf{X}}, P_{\mathrm{mix}}^B \times P^E \times P^{\mathbf{X}}) \quad (58)$$

holds, where B is the random variable $f_{\mathbf{X}}(A)$. Hence, when the expectation $\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A),E}|E)$ is sufficiently small, the random variable $f_{\mathbf{X}}(A)$ is almost independent of the random variables $\mathbf{X}$ and E. So, the above value is suitable even when we randomly choose a hash function.

In order to evaluate the average performance, we define the quantity

$$\begin{aligned}\phi(t|A|E|P^{A,E}) &:= \log \sum_e P^E(e) (\sum_a P^{A|E}(a|e)^{\frac{1}{1-t}})^{1-t} \\ &= \log \sum_e (\sum_a P^{A,E}(a,e)^{\frac{1}{1-t}})^{1-t}. \end{aligned}$$

Note that when Eve's random variable E takes a continuous value in the set $\mathcal{E}$, the relation (59) holds by defining $\phi(t|A|E|P^{A,E})$ in the following way.

$$\phi(t|A|E|P^{A,E}) := \log \int_{\mathcal{E}} P^E(e) de (\sum_a P^{A|E}(a|e)^{\frac{1}{1-t}})^{1-t}.$$

This definition does not depend on the choice of the measure on $\mathcal{Y}$.

By using Theorem 1 and putting $t = \frac{s}{1+s}$, any universal$_2$ hash functions $\{f_{\mathbf{X}}\}$ satisfies the inequality:

$$\begin{aligned}\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A),E}|E) &\le 3M^{\frac{s}{1+s}} \mathrm{E}_e (\sum_a P^{A|E}(a|e)^{1+s})^{\frac{1}{1+s}} \\ &= 3M^t e^{\phi(t|A|E|P^{A,E})} \end{aligned} \quad (59)$$

for $0 \le t \le \frac{1}{2}$. Therefore, there exists a function f such that

$$\begin{aligned}d_1(P^{f(A),E}|E) &\le 3M^{\frac{s}{1+s}} \mathrm{E}_e (\sum_a P^{A|E}(a|e)^{1+s})^{\frac{1}{1+s}} \\ &= 3M^t e^{\phi(t|P^{A,E})} \end{aligned} \quad (60)$$

Next, we consider the case when our distribution $P^{A_nE_n}$ is given by the n-fold independent and identical distribution of P^{AE}, i.e, $(P^{A,E})^n$. Ahlswede and Csiszár [7] showed that the optimal generation rate

$$\begin{aligned}&G(P^{A,E}) \\ :=& \sup_{\{(f_n,M_n)\}} \left\{ \lim_{n\to\infty} \frac{\log M_n}{n} \,\middle|\, \lim_{n\to\infty} d_1(P^{f_n(A_n),E_n}|E_n) = 0 \right\} \end{aligned}$$

equals the conditional entropy $H(A|E)$. That is, the generation rate $R = \lim_{n\to\infty} \frac{\log M_n}{n}$ is smaller than $H(A|E)$. The quantity $d_1(P^{f_n(A_n),E_n}|E_n)$ goes to zero. In order to treat the speed of this convergence, we focus on the supremum of the *exponential rate of decrease (exponent)* for $d_1(P^{f_n(A_n),E_n}|E_n)$ for a given R

$$\begin{aligned}&e_1(P^{A,E}|R) \\ :=& \sup_{\{(f_n,M_n)\}} \Big\{ \lim_{n\to\infty} \frac{-1}{n} \log d_1(P^{f_n(A_n),E_n}|E_n) \Big| \\ & \qquad \lim_{n\to\infty} \frac{-1}{n} \log M_n \le R \Big\}. \end{aligned}$$

Since the relation $\phi(t|A^n|E^n|(P^{A,E})^n) = n\phi(t|A|E|P^{A,E})$ holds, the inequality (60) implies that

$$e_1(P^{A,E}|R) \geq -\phi(t|A|E|P^{A,E}) - tR. \tag{61}$$

for $t \in [0, 1/2]$. That is, taking the maximum concerning $t \in [0, 1/2]$, we obtain

$$e_1(P^{A,E}|R) \geq e_\phi(A|E|P^{A,E}|R), \tag{62}$$

where

$$\begin{aligned} e_\phi(A|E|P^{A,E}|R) &:= \max_{0\leq t\leq \frac{1}{2}} -\phi(t|A|E|P^{A,E}) - tR \\ &= \max_{0\leq s\leq 1} -\phi(\frac{s}{1+s}|A|E|P^{A,E}) - \frac{s}{1+s}R. \end{aligned}$$

Since $\frac{d}{dt}\phi(t|P^{A,E})\big|_{t=0} = \frac{d(sH_{1+s}(A|E|P^{A,E})}{ds}\Big|_{s=0} = -H(A|E)$, the right hand sides of (62) and (63) are strictly greater than 1 for $R < H(A|E)$.

B. Comparison with the previous paper [6]

Next, we show how better our bound is than that by the previous paper [6]. The previous paper [6] shows the following in Section IIA. There exists a sequence of functions $f_n : \mathcal{A}^n \to \{1, \ldots, \lfloor e^{nR} \rfloor\}$ such that

$$\begin{aligned} &\lim_{n\to\infty} \frac{-1}{n} \log D(P^{f_n(A_n),E_n} \| P_{\mathrm{mix}}^{f_n(A_n)} \times P^{E_n}) \\ \geq &\max_{0\leq s\leq 1} sH_{1+s}(A|E|P^{A,E}) - sR, \end{aligned}$$

where we define the function

$$\begin{aligned} sH_{1+s}(A|E|P^{A,E}) &:= -\log \sum_{a,e} P^E(e) P^{A|E}(a|e)^{1+s} \\ &= -\log \sum_{a,e} P^{A,E}(a,e)^{1+s} P^E(e)^{-s} \end{aligned}$$

for $s \in [0, 1]$. Hence, applying Pinsker inequality (6), we obtain

$$\begin{aligned} e_1(P^{A,E}|R) \geq &\lim_{n\to\infty} \frac{-1}{n} \log d_1(P^{f_n(A_n),E_n}|E_n) \\ \geq &\tilde{e}_H(A|E|P^{A,E}|R) \end{aligned} \tag{63}$$

where

$$\begin{aligned} \tilde{e}_H(A|E|P^{A,E}|R) &:= \max_{0\leq s\leq 1} \frac{sH_{1+s}(A|E|P^{A,E}) - sR}{2} \\ &= \max_{0\leq t\leq \frac{1}{2}} \frac{tH_{\frac{1}{1-t}}(A|E|P^{A,E}) - tR}{2-2t} \end{aligned}$$

with $s = \frac{t}{1-t}$. Concerning the comparison of both bounds, we prepare the following lemma.

Lemma 7: The inequality

$$-\frac{s}{1+s} H_{1+s}(A|E|P^{A,E}) \geq \phi(\frac{s}{1+s}|A|E|P^{A,E}) \tag{64}$$

holds for $s \in (0, \infty)$. The equality holds if and only if the Rényi entropy $H_{1+s}(A|P^{A|E=e})$ does not depends on the choice e at the support of P^E.

Proof: Then, applying Jensen inequality to the concave function $x \mapsto x^{\frac{1}{1+s}}$, we have

$$\begin{aligned} &e^{-\frac{sH_{1+s}(A|E|P^{A,E})}{1+s}} = (\sum_e P^E(e) \sum_a P^{A|E}(a|e)^{1+s})^{\frac{1}{1+s}} \\ \geq &\sum_e P^E(e) (\sum_a P^{A|E}(a|e)^{1+s})^{\frac{1}{1+s}} = e^{\phi(\frac{s}{1+s}|A|E|P^{A,E})}. \end{aligned}$$

Thus, the equality condition is that the value $\sum_a P^{A|E}(a|e)^{1+s}$ does not depends on the choice e at the support of P^E. Hence, we obtain the desired argument. ∎

In order to compare two bounds $e_\phi(A|E|P^{A,E}|R)$ and $\tilde{e}_H(A|E|P^{A,E}|R)$, we introduce the following value:

$$\begin{aligned} e_H(A|E|P^{A,E}|R) &:= \max_{0\leq s\leq 1} \frac{sH_{1+s}(A|E|P^{A,E}) - sR}{1+s} \\ &= \max_{0\leq t\leq \frac{1}{2}} tH_{\frac{1}{1-t}}(A|E|P^{A,E}) - tR \end{aligned}$$

Then, we obtain the following lemma.

Lemma 8:

$$e_\phi(A|E|P^{A,E}|R) \geq e_H(A|E|P^{A,E}|R) \geq \tilde{e}_H(A|E|P^{A,E}|R) \tag{65}$$

for $R < H(A|E)$. The equality in the first inequality holds if and only if the Rényi entropy $H_{1+s_0}(A|P^{A|E=e})$ does not depends on the choice e at the support of P^E for $s_0 := \mathrm{argmax}_{0\leq s\leq 1} -\phi(\frac{s}{1+s}|A|E|P^{A,E}) - \frac{s}{1+s}R$. The equality in the second inequality holds if and only if $\frac{sH_2(A|E|P^{A,E})-R}{2} = \max_{0\leq s\leq 1} \frac{sH_{1+s}(A|E|P^{A,E})-sR}{1+s}$.

Therefore, our exponent $e_\phi(A|E|P^{A,E}|R)$ is strictly better than the exponent $\tilde{e}_H(A|E|P^{A,E}|R)$ by [6, Section IIA] except for the case satisfying the following two conditions: (i) $-\phi(\frac{1}{2}|A|E|P^{A,E}) - \frac{1}{2}R = \max_{0\leq s\leq 1} -\phi(\frac{s}{1+s}|A|E|P^{A,E}) - \frac{s}{1+s}R$. (ii) $H_2(A|P^{A|E=e})$ does not depends on the choice e at the support of P^E.

For example, we consider the following case: $\mathcal{A}$ equals $\mathcal{E}$, the set $\mathcal{A}$ has the module structure, (i.e., $\mathcal{A}$ is an Abelian group) and the conditional distribution $P^{A|E}(a|e)$ has the form $P^A(a-e)$. Then, the equality condition for the first inequality holds. Since

$$\begin{aligned} &e^{\phi(\frac{s}{1+s}|A|E|P^{A,E})} = \sum_e P^E(e) (\sum_a P^{A|E}(a|e)^{1+s})^{\frac{1}{1+s}} \\ = &\sum_e P^E(e) e^{-\frac{sH_{1+s}(A|P^A)}{1+s}} = e^{-\frac{sH_{1+s}(A|P^A)}{1+s}}. \end{aligned}$$

and

$$\begin{aligned} &e^{-sH_{1+s}(A|E|P^{A,E})} = \sum_e P^E(e) \sum_a P^{A|E}(a|e)^{1+s} \\ = &\sum_e P^E(e) \sum_a P^A(a-e)^{1+s} \\ = &\sum_e P^E(e) e^{-\frac{sH_{1+s}(A|P^A)}{1+s}} = e^{-\frac{sH_{1+s}(A|P^A)}{1+s}}, \end{aligned}$$

bounds $e_\phi(A|E|P^{A,E}|R)$ and $\tilde{e}_H(A|E|P^{A,E}|R)$ can be simplified to

$$\begin{aligned} e_\phi(A|E|P^{A,E}|R) &= e_H(A|E|P^{A,E}|R) = e_H(A|P^A|R) \\ \tilde{e}_H(A|E|P^{A,E}|R) &= \tilde{e}_H(A|P^A|R), \end{aligned}$$

where

$$\begin{aligned} e_H(A|P^A|R) &:= \max_{0\le s\le 1} \frac{sH_{1+s}(A|P^A)-sR}{1+s} \\ &= \max_{0\le t\le 1/2} tH_{\frac{1}{1-t}}(A|P^A)-tR \\ \tilde{e}_H(A|P^A|R) &:= \max_{0\le s\le 1} \frac{sH_{1+s}(A|P^A)-sR}{2} \\ &= \max_{0\le t\le 1/2} \frac{tH_{\frac{1}{1-t}}(A|P^A)-tR}{2-2t}. \end{aligned}$$

In particular, the both exponents are numerically plotted in Fig. 2 when $\mathcal{A}=\{0,1\}$, and $P^A(0)=p$, $P^A(1)=1-p$.

Proof: The first inequality and its equality condition follow from Lemma 7 and the definitions of $e_\phi(P^{A,E}|R)$ and $e_H(P^{A,E}|R)$. The second inequality follows from the inequality $\frac{1}{2}\le\frac{1}{1+s}$ for $s\in[0,1]$. Since the equality holds only when $s=1$, we obtain the equality condition for the second inequality. ∎

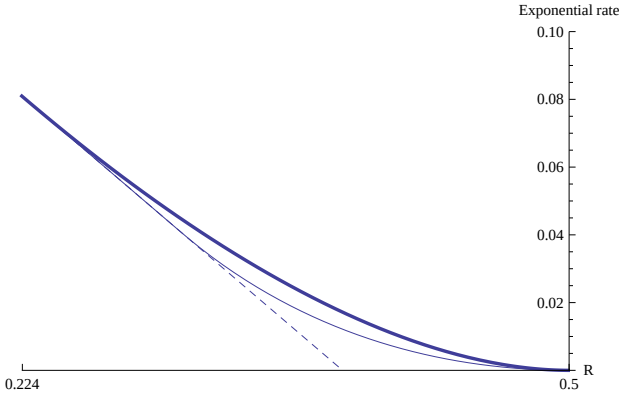

Fig. 2. Lower bounds of $e_1(P^{AE}|R)$. Thick line: $e_H(A|P^A|R)$ (The present paper), Normal line: $\tilde{e}_H(A|P^A|R)$ by [6]), Dashed line: $\frac{H_2(A|P^A)-sR}{2}$ (direct application of (11) without smoothing) $p=0.200$, $h(p)=H(A)=0.500$, $2\frac{d(sH_{1+s}(A))}{ds}|_{s=1}-H_2(A)=0.224$.

VI. The wire-tap channel in a general framework

Next, we consider the wire-tap channel model, in which the eavesdropper (wire-tapper) Eve and the authorized receiver Bob receive the information from the authorized sender Alice. In this case, in order for Eve to have less information, Alice chooses a suitable encoding. This problem is formulated as follows. Let $\mathcal{Y}$ and $\mathcal{Z}$ be the probability spaces of Bob and Eve, and $\mathcal{X}$ be the set of alphabets sent by Alice. Then, the main channel from Alice to Bob is described by $W^B: x\mapsto W_x^B$, and the wire-tapper channel from Alice to Eve is described by $W^E: x\mapsto W_x^E$. That is, W_x^B is the output distribution on the Bob's side with Alice's input x, and W_x^E is the output distribution on the Eve's side with Alice's input x. In this setting, in order to send the secret message in $\{1,\ldots,M\}$ subject to the uniform distribution, Alice chooses M distributions $Q_1,\ldots,Q_M$ on $\mathcal{X}$, and she generates $x\in\mathcal{X}$ subject to Q_i when she wants to send the message $i\in\{1,\ldots,M\}$. Bob prepares M disjoint subsets $\mathcal{D}_1,\ldots,\mathcal{D}_M$ of $\mathcal{Y}$ and judges that a message is i if y belongs to $\mathcal{D}_i$. Therefore, the triplet $(M,\{Q_1,\ldots,Q_M\},\{\mathcal{D}_1,\ldots,\mathcal{D}_M\})$ is called a code, and is described by Φ. Its performance is given by the following three quantities. The first is the size M, which is denoted by $|\Phi|$. The second is the average error probability $\epsilon_B(\Phi)$:

$$\epsilon_B(\Phi)\stackrel{\rm def}{=}\frac{1}{M}\sum_{i=1}^M W_{Q_i}^B(\mathcal{D}_i^c),$$

and the third is Eve's distinguishability $d_1(\Phi|E)$:

$$\begin{aligned} d_1(\Phi|E) :=& d_1(W_\Phi^E\times P_{\rm mix}^M, W^E[\Phi]) \\ W_\Phi^E(e) :=& \sum_i \frac{1}{M}W_{Q_i}^E(e), \quad W^E[\Phi](i,e) := \frac{1}{M}W_{Q_i}^E(e). \end{aligned}$$

The quantity $d_1(\Phi|E)$ gives an upper bound for the probability that Eve can succeed in distinguishing whether Alice's information belongs to a given subset. So, the value can be regarded as Eve's distinguishability. In order to calculate these values, we introduce the following quantity.

$$\phi(t|W,p):=\log\sum_y\left(\sum_x p(x)(W_x(y))^{1/(1-t)}\right)^{1-t}.$$

When the random variable Y takes a continuous value in the set $\mathcal{Y}$ while X takes discrete value, the above definition can be changed to

$$\phi(t|W,p):=\log\int_{\mathcal{Y}}\left(\sum_x p(x)(W_x(y))^{1/(1-t)}\right)^{1-t}dy$$

This definition does not depend on the choice of the measure on $\mathcal{Y}$. That is, when $\tilde{W}_x(y)f(y)=W_x(y)$ for a positive function f,

$$\phi(t|W,p)=\log\int_{\mathcal{Y}}\left(\sum_x p(x)(\tilde{W}_x(y))^{1/(1-t)}\right)^{1-t}f(y)dy.$$

As is shown as Lemma 1 of [6], $\phi(t|W,p)$ satisfies the following lemma.

Lemma 9: The function $p\mapsto e^{\phi(t|W,p)}$ is convex for $t\in[-1,0]$, and is concave for $t\in[0,1]$.

Now, using the function $\phi(t)$, we make a code for the wire-tap channel based on the random coding method. For this purpose, we make a protocol to share a random number. First, we generate the random code $\Phi(\mathbf{Y})$ with size LM, which is described as $\Phi(\mathbf{Y})(a)=Y_a$ for $a=1,\ldots,LM$ by using the LM independent and identical random variables $\mathbf{Y}=(Y_1,\ldots,Y_{ML})$ subject to the distribution p on $\mathcal{X}$. Gallager [20] showed that the ensemble expectation of the average error probability concerning decoding the input message A is less than $(ML)^t e^{\phi(-t|W^B,p)}$ for $0\le t\le 1$ when Bob applies the maximum likelihood decoder $\mathcal{D}'(\mathbf{Y})$ of the code $\Phi(\mathbf{Y})$. After sending the random variable A taking values in the set with the cardinality ML, Alice and Bob apply the above universal$_2$ hash functions $f_{\mathbf{X}}$ to the random variable A and generate another piece of data of size M. Here, we assume that the ensemble $\{f_{\mathbf{X}}\}$ satisfies Condition 2. Then, Alice and Bob share the random variable $f_{\mathbf{X}}(A)$ with size M. This protocol is denoted by $\Phi(\mathbf{X},\mathbf{Y})'$

Let E be the random variable of the output of Eve's channel W^E. When p is the uniform distribution on the set $\mathcal{C} := \{1, \ldots, ML\}$ and the joint distribution $P^{C,E}$ is given by $P^{C,E}(c, e) := p(c)W_c^E(e)$, the equations

$$\begin{aligned} e^{\phi(t|P^{C,E})} =& \frac{1}{M^t L^t} \sum_e \left(\sum_a p(c)(W_c^E(e))^{\frac{1}{1-t}} \right)^{1-t} \\ =& \frac{e^{\phi(t|W^E,p)}}{M^t L^t}. \end{aligned} \tag{66}$$

hold.

For a given code $\Phi(\mathbf{Y})$, we apply the inequality (59) to Eve's distinguishability. Then,

$$\mathrm{E}_{\mathbf{X}|\mathbf{Y}} d_1(\Phi(\mathbf{X}, \mathbf{Y})'|E) \le 3 \frac{e^{\phi(t|W^E, p_{\mathrm{mix},\Phi(\mathbf{Y})})}}{L^t} \tag{67}$$

for $0 \le \forall t \le \frac{1}{2}$. The concavity of $e^{\phi(t|W^E,p)}$ (Lemma 9) guarantees that

$$\begin{aligned} &\mathrm{E}_{\mathbf{X},\mathbf{Y}} d_1(\Phi(\mathbf{X}, \mathbf{Y})'|E) \le 3\mathrm{E}_{\mathbf{Y}} \frac{e^{\phi(t|W^E, p_{\mathrm{mix},\Phi(\mathbf{Y})})}}{L^t} \\ \le& 3 \frac{e^{\phi(t|W^E,p)}}{L^t} \end{aligned}$$

for $0 \le \forall t \le \frac{1}{2}$.

Now, we make a code for wire-tap channel by modifying the above protocol $\Phi(\mathbf{X}, \mathbf{Y})'$. First, we choose the distribution Q_i to be the uniform distribution on $f_{\mathbf{X}}^{-1}\{i\}$. When Alice wants to send the secret message i, before sending the random variable A, Alice generates the random number A subject to the distribution Q_i. Alice sends the random variable A. Bob recovers the random variable A by using the maximum likelihood decoder $\mathcal{D}'(\mathbf{Y})$, and applies the function $f_{\mathbf{X}}$. Then, Bob decodes Alice's message i, and this code for wire-tap channel W^B, W^E is denoted by $\Phi(\mathbf{X}, \mathbf{Y})$. Since the ensemble $\{f_{\mathbf{X}}\}$ satisfies Condition 2 and the secret message i obeys the uniform distribution on $\{1, \ldots, M\}$, this protocol $\Phi(\mathbf{X}, \mathbf{Y})$ has the same performance as the above protocol $\Phi(\mathbf{X}, \mathbf{Y})'$.

Finally, we consider what code is derived from the above random coding discussion. Using the Markov inequality, we obtain

$$\begin{aligned} &\mathrm{P}_{\mathbf{X},\mathbf{Y}}\{\epsilon_B(\Phi(\mathbf{X}, \mathbf{Y})) \le 3\mathrm{E}_{\mathbf{X},\mathbf{Y}} \epsilon_B(\Phi(\mathbf{X}, \mathbf{Y}))\} \ge \frac{2}{3} \\ &\mathrm{P}_{\mathbf{X},\mathbf{Y}}\{d_1(\Phi(\mathbf{X}, \mathbf{Y})|E) \le 3\mathrm{E}_{\mathbf{X},\mathbf{Y}} d_1(\Phi(\mathbf{X}, \mathbf{Y})|E)\} \ge \frac{2}{3}. \end{aligned}$$

Therefore, the existence of a good code is guaranteed in the following way. That is, we give the concrete performance of a code whose existence is shown in the above random coding method.

Theorem 5: There exists a code Φ for any integers L, M, and any probability distribution p on $\mathcal{X}$ such that $|\Phi| = M$ and

$$\begin{aligned} \epsilon_B(\Phi) \le& 3 \min_{0 \le t \le 1} (ML)^t e^{\phi(-t|W^B,p)}, \\ d_1(\Phi|E) \le& 9 \min_{0 \le t \le \frac{1}{2}} \frac{e^{\phi(t|W^E,p)}}{L^t}. \end{aligned}$$

In the n-fold discrete memoryless channels W^{B_n} and W^{E_n} of the channels W^B and W^E, the additive equation $\phi(t|W^{B_n}, p) = n\phi(t|W^B, p)$ holds. Thus, there exists a code Φ_n for any integers L_n, M_n, and any probability distribution p on $\mathcal{X}$ such that $|\Phi_n| = M_n$ and

$$\begin{aligned} \epsilon_B(\Phi) \le& 3 \min_{0 \le t \le 1} (M_n L_n)^t e^{n\phi(-t|W^B,p)}, \\ d_1(\Phi_n|E) \le& 9 \min_{0 \le t \le \frac{1}{2}} \frac{e^{n\phi(t|W^E,p)}}{L_n^t}. \end{aligned}$$

Since $\lim_{t \to 0} \frac{\phi(t|W^E,p)}{t} = I(p : W^E)$, the rate $\max_p I(p : W^B) - I(p : W^E)$ can be asymptotically attained. Therefore, when the sacrifice information rate is R, i.e., $L_n \cong e^{nR}$, the exponential rate of decrease for Eve's distinguishability is greater than

$$e_\phi(R|W^E, p) := \max_{0 \le t \le 1/2} tR - \phi(t|W^E, p).$$

VII. Comparison with existing bound

In Subsection VII-A, we compare our exponent $e_\phi(R|W^E, p)$ with those derived by [17], [6] in the general setting. In Subsections VII-B and VII-C, using discussion in Subsection V-B, we treat this comparison in special cases more deeply.

A. General case

Now, we compare the lower obtained bound $e_\phi(R|W^E, p)$ for the exponential rate of decrease for Eve's distinguishability with existing lower bounds [17], [6]. Using the quantity

$$\begin{aligned} \psi(t|W, p) :=& \log \sum_y \left(\sum_x p(x)(W_x(y))^{1+t} \right) W_p(y)^{-t} \\ W_p(y) :=& \sum_x p(x) W_x(y), \end{aligned} \tag{68}$$

the previous paper [17] derived the following lower bound of this exponential rate of decrease:

$$\begin{aligned} e_\psi(R|W^E, p) :=& \max_{0 \le s \le 1} \frac{sR - \psi(s|W^E, p)}{1+s} \\ =& \max_{0 \le t \le 1/2} tR - (1-t)\psi(\frac{t}{1-t}|W^E, p). \end{aligned} \tag{69}$$

The other previous paper [6] also derived the following lower bound:

$$\max_{0 \le s \le 1} sR - \psi(s|W^E, p) \tag{70}$$

for the exponential rate of decrease for the mutual information. By applying a discussion similar to Subsection V-B and Pinsker inequality (9), the bound (70) yields the bound

$$\tilde{e}_\psi(R|W^E, p) := \max_{0 \le s \le 1} \frac{sR - \psi(s|W^E, p)}{2}, \tag{71}$$

which is smaller than the lower bound $e_\psi(R|W^E, p)$ because $\frac{1}{2} \le \frac{1}{1+s}$ for $0 \le s \le 1$. Hence, in order to show the superiority of our bound $e_\phi(R|W^E, p)$, it is sufficient to show the superiority over the bound $e_\psi(R|W^E, p)$.

In the following, we compare the two bounds $e_\phi(R|W^E, p)$ and $e_\psi(R|W^E, p)$ For this purpose, we treat $e^{\phi(t|W^E,p)}$ and

$e^{(1-t)\psi(\frac{t}{1-t}|W^E,p)}$ for $0 \le t \le \frac{1}{2}$. Reverse Hölder inequality [28] with the measurable space $(\mathcal{X}, p)$ is given as

$$\begin{aligned}&\sum_{x\in\mathcal{X}} p(x)|X(x)Y(x)| \\ \ge&(\sum_{x\in\mathcal{X}} p(x)|X(x)|^{\frac{1}{1+s}})^{1+s}(\sum_{x\in\mathcal{X}} p(x)|Y(x)|^{-\frac{1}{s}})^{-s}\end{aligned}$$

for $s \ge 0$. Using this inequality, we obtain

$$\begin{aligned}&\sum_y \left[\sum_x p(x)(W_x(y))^{1+s}\right] W_p(y)^{-s} \\ \ge& \left(\sum_y \left[\sum_x p(x)(W_x(y))^{1+s}\right]^{\frac{1}{1+s}}\right)^{1+s} \cdot \left(\sum_y W_p(y)^{-s\cdot-\frac{1}{s}}\right)^{-s} \\ =& \left(\sum_y \left[\sum_x p(x)(W_x(y))^{1+s}\right]^{\frac{1}{1+s}}\right)^{1+s}.\end{aligned}$$

Substituting $s = \frac{t}{1-t}$, we obtain

$$\begin{aligned}&\sum_y \left[\sum_x p(x)(W_x(y))^{\frac{1}{1-t}}\right] W_p(y)^{\frac{-t}{1-t}} \\ \ge& \left(\sum_y \left[\sum_x p(x)(W_x(y))^{\frac{1}{1-t}}\right]^{1-t}\right)^{\frac{1}{1-t}},\end{aligned}$$

which implies

$$\begin{aligned}&e^{(1-t)\psi(\frac{t}{1-t}|W^E,p)} \\ =& \left(\sum_y \left[\sum_x p(x)(W_x(y))^{\frac{1}{1-t}}\right] W_p(y)^{\frac{-t}{1-t}}\right)^{1-t} \\ \ge& \sum_y \left[\sum_x p(x)(W_x(y))^{\frac{1}{1-t}}\right]^{1-t} = e^{\phi(t|W^E,p)}.\end{aligned}$$

Thus, our bound $e_\phi(R|W^E, p)$ for the exponential rate of decrease is better than the existing bound $e_\psi(R|W^E, p)$ [17].

Example 1: Assume that $\mathcal{X} = \mathcal{E} = \{0, 1\}$. We consider the following channel.

$$W_0(0) = a,\ W_0(1) = 1-a,\ W_1(0) = 1-9a,\ W_1(1) = 9a.$$

When $p(0) = 1/2, p(1) = 1/2$,

$$\begin{aligned}I(p, W) =& h(1/2 - 5p) - \frac{(h(p) + h(9p)}{2} \\ \psi(t|p, W) =& \log\left((\frac{a^{1+t} + (1-9a)^{1+t}}{2}(\frac{1}{2} - 5p)^{-t}\right. \\ &\left. + (\frac{(9p)^{1+t} + (1-p)^{1+t}}{2}(1/2 + 5p)^{-t})\right) \\ \phi(t|p, W) =& \log\left((\frac{p^{1/(1-t)} + (1-9p)^{1/(1-t)}}{2})^{1-t}\right. \\ &\left. + (\frac{(9p)^{1/(1-t)} + (1-p)^{1/(1-t)}}{2})^{1-t}\right).\end{aligned}$$

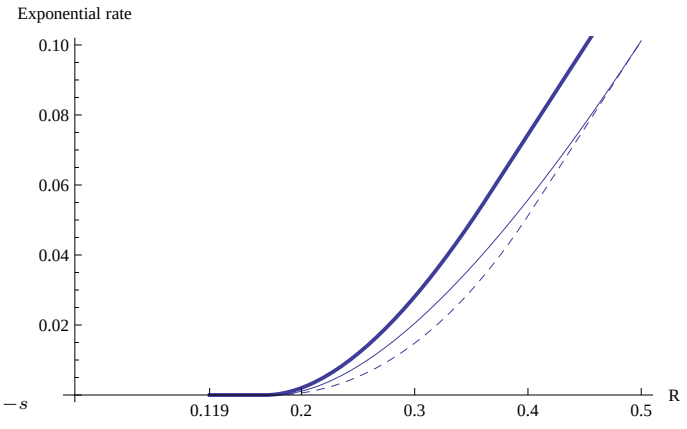

Fig. 3. Lower bounds of exponent. Thick line: $e_\phi(R|W, p)$ (The present paper), Normal line: $e_\psi(R|W, p)$ [17], Dashed line: $\tilde{e}_\psi(R|W, p)$ [6] $a = 0.0500$, $I(p, W) = 0.119$.

Then, the three bounds $e_\phi(R|W, p)$, $e_\psi(R|W, p)$, and $\tilde{e}_\psi(R|W, p)$ with $a = 0.05$ are numerically compared as in Fig. 3.

B. Additive case

Next, we consider a more specific case. When $\mathcal{X} = \mathcal{Z}$ and $\mathcal{X}$ is a module and $W_x(z) = W_0(z - x) = P^X(z - x)$, the channel W is called *additive*.

Since

$$\begin{aligned}&e^{(1-t)\psi(\frac{t}{1-t}|W^E,p_{\rm mix})} = e^{\phi(t|W^E,p_{\rm mix})} \\ =&|\mathcal{X}|^t e^{-tH_{\frac{1}{1-t}}(X|P^X)},\end{aligned} \tag{72}$$

any additive channel W^E satisfies

$$\begin{aligned}&e_\psi(R|W^E, p_{\rm mix}) = e_\phi(R|W^E, p_{\rm mix}) \\ =& \max_{0\le t\le\frac{1}{2}} t(R - \log|\mathcal{X}|) + tH_{\frac{1}{1-t}}(X|P^X)) \\ =&e_H(X|P^X|\log|\mathcal{X}| - R)\end{aligned} \tag{73}$$

and

$$\begin{aligned}&\tilde{e}_\psi(R|W^E, p_{\rm mix}) = \max_{0\le t\le\frac{1}{2}} \frac{t(R - \log|\mathcal{X}|) + tH_{\frac{1}{1-t}}(X|P^X)}{2-2t} \\ =&\tilde{e}_H(X|P^X|\log|\mathcal{X}| - R)\end{aligned}$$

for the uniform distribution $p_{\rm mix}$ on $\mathcal{X}$.

Hence, our bound $e_\phi(R|W^E, p_{\rm mix})$ is the same as the previous bound $e_\psi(R|W^E, p_{\rm mix})$. However, since $\frac{1}{2-2t} < 1$ for $t \in [0, 1/2)$, our bound $e_\phi(R|W^E, p_{\rm mix})$ is strictly better than the bound $\tilde{e}_\psi(R|W^E, p_{\rm mix})$ by the other previous paper [6] when the maximum is attained by $t \in [0, 1/2)$.

C. General additive case

We consider a more general case. Eve is assumed to have two random variables $Z \in \mathcal{X}$ and $Z' \in \mathcal{Z}'$. The first random variable Z is the output of an additive channel depending on the second variable Z'. That is, the channel $W^E_x(z, z')$ can be written as $W^E_x(z, z') = P^{X,Z'}(z - x, z')$, where $P^{X,Z'}$ is a joint distribution. Hereinafter, this channel model is called a

general additive channel. This channel is also called a regular channel[21]. For this channel model, we obtain

$$\begin{aligned} e^{\phi(s|W^E,P_{\mathrm{mix},\mathcal{X}})} =& \sum_{z,z'}(\sum_x \frac{1}{|\mathcal{X}|} W_x^E(z,z')^{\frac{1}{1-s}})^{1-s} \\ =& \sum_{z,z'}(\sum_x \frac{1}{|\mathcal{X}|} P^{X,Z'}(z-x,z')^{\frac{1}{1-s}})^{1-s} \\ =& \frac{1}{|\mathcal{X}|^{1-s}} \sum_{z,z'}(\sum_x P^{X,Z'}(-x,z')^{\frac{1}{1-s}})^{1-s} \\ =& \frac{|\mathcal{X}|}{|\mathcal{X}|^{1-s}} \sum_{z'}(\sum_x P^{X,Z'}(x,z')^{\frac{1}{1-s}})^{1-s} \\ =& |\mathcal{X}|^s e^{\phi(s|X|Z'|P^{X,Z'})}. \end{aligned} \tag{74}$$

and

$$\begin{aligned} & e^{\psi(s|W^E,p_{\mathrm{mix}})} \\ =& \sum_{z,z'}(\sum_x \frac{1}{|\mathcal{X}|} W_x^E(z,z')^{1+s})(\sum_x \frac{1}{|\mathcal{X}|} W_x^E(z,z'))^{-s} \\ =& |\mathcal{X}|^{s-1} \sum_{z,z'}(\sum_x P^{X,Z'}(z-x,z')^{1+s})(\sum_x P^{X,Z'}(z-x,z'))^{-s} \\ =& |\mathcal{X}|^{s-1} \sum_{z,z'}(\sum_x P^{X,Z'}(-x,z')^{1+s}) P^{Z'}(z')^{-s} \\ =& |\mathcal{X}|^{s-1}|\mathcal{X}| \sum_{z'}\sum_x P^{X,Z'}(x,z')^{1+s} P^{Z'}(z')^{-s} \\ =& |\mathcal{X}|^s e^{-sH_{1+s}(X|Z'|P^{X,Z'})}. \end{aligned} \tag{75}$$

Then, the equalities

$$\begin{aligned} & e_\phi(R|W^E,p_{\mathrm{mix}}) \\ =& \max_{0\le t\le \frac{1}{2}} t(R-\log|\mathcal{X}|) - \phi(t|X|Z'|P^{X,Z'}) \\ =& e_\phi(X|Z'|P^{X,Z'}|\log|\mathcal{X}|-R) \end{aligned} \tag{76}$$

$$\begin{aligned} & e_\psi(R|W^E,p_{\mathrm{mix}}) \\ =& \max_{0\le t\le \frac{1}{2}} t(R-\log|\mathcal{X}|) + tH_{\frac{1}{1-t}}(X|Z'|P^{X,Z'}) \\ =& e_H(X|Z'|P^{X,Z'}|\log|\mathcal{X}|-R) \end{aligned} \tag{77}$$

$$\begin{aligned} & \tilde{e}_\psi(R|W^E,p_{\mathrm{mix}}) \\ =& \max_{0\le t\le \frac{1}{2}} \frac{t(R-\log|\mathcal{X}|) + tH_{\frac{1}{1-t}}(X|Z'|P^{X,Z'})}{2-2t} \\ =& \tilde{e}_H(X|Z'|P^{X,Z'}|\log|\mathcal{X}|-R) \end{aligned} \tag{78}$$

hold.

Hence, the observation in Section V-B can be applied to the comparison among $e_\phi(R|W^E,p_{\mathrm{mix}})$, $e_\psi(R|W^E,p_{\mathrm{mix}})$, and $\tilde{e}_\psi(R|W^E,p_{\mathrm{mix}})$. Due to Lemma 8, $e_\phi(R|W^E,p_{\mathrm{mix}})$ is strictly better than $e_\psi(R|W^E,p_{\mathrm{mix}})$ and $\tilde{e}_\psi(R|W^E,p_{\mathrm{mix}})$ except for the special case mentioned in Lemma 8.

VIII. Wire-tap channel with linear coding

In a practical sense, we need to take into account the decoding time. For this purpose, we often restrict our codes to linear codes. In the following, we consider the case where the sender's space $\mathcal{X}$ has the structure of a module. When an error correcting code is given as a submodule $C_1 \subset \mathcal{X}$ and the decoder by the authorized receiver is given as $\{\mathcal{D}_x\}_{x\in C_1}$, our code for a wire-tap channel is given as $\Phi_{C_1,C_2} = (|C_1/C_2|, \{Q_{[x]}\}_{[x]\in C_1/C_2}, \{\mathcal{D}_{[x]}\}_{[x]\in C_1/C_2})$ based on a submodule C_2 of C_1 as follows. The encoding $Q_{[x]}$ is given as the uniform distribution on the coset $[x] := x + C_2$, and the decoding $\mathcal{D}_{[x]}$ is given as the subset $\cup_{x'\in x+C_2}\mathcal{D}_{x'}$. Next, we consider a submodule $C_2(\mathbf{X})$ of C_1 with cardinality $|C_2(\mathbf{X})| = L$ that is labeled by a random variable $\mathbf{X}$. Then, the module $C_2(\mathbf{X})$ can be regarded as a random variable. Now, we impose the module $C_2(\mathbf{X})$ the following condition.

Condition 4: Any element $x \neq 0 \in C_1$ is included in $C_2(\mathbf{X})$ with probability at most $\frac{L}{|C_1|}$.

Then, using (67), we can evaluate the performance of the constructed code in the following way.

Theorem 6: Choose the subcode $C_2(\mathbf{X})$ according to Condition 4. We construct the code $\Phi_{C_1,C_2(\mathbf{X})}$ by choosing the distribution $Q_{[x]}$ to be the uniform distribution on $[x]$ for $[x] \in C_1/C_2(\mathbf{X})$. Then, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3\frac{e^{\phi(t|W^E,P_{\mathrm{mix},C_1})}}{L^t} \quad 0 \le \forall t \le \frac{1}{2}, \tag{79}$$

where $P_{\mathrm{mix},S}$ is the uniform distribution on the subset S.

When the channel W^E is additive, i.e., $W_x^E(z) = P^X(z-x)$, the equation $\phi(t|W^E,P_{\mathrm{mix},C_1+x}) = \phi(t|W^E,P_{\mathrm{mix},C_1})$ holds for any x. Thus, the concavity of $e^{\phi(t|W^E,p)}$ (Lemma 9) implies that

$$\phi(t|W^E,P_{\mathrm{mix},C_1}) \le \phi(t|W^E,P_{\mathrm{mix},\mathcal{X}}). \tag{80}$$

Thus, combining (79), (80), and (72), we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3\frac{|\mathcal{X}|^t e^{-tH_{\frac{1}{1-t}}(X|P)}}{L^t} \tag{81}$$

for $0 < \forall t < \frac{1}{2}$. That is, when $L = e^R$, taking the minimum concerning $0 < \forall t < \frac{1}{2}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3e^{-e_H(X|P^X|\log|\mathcal{X}|-R)}. \tag{82}$$

When the additive noise obeys the n-fold i.i.d. of P on $\mathcal{X}^n$ and $L = e^{nR}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3e^{-ne_H(X|P^X|\log|\mathcal{X}|-R)}. \tag{83}$$

Similarly, when the channel W^E is general additive, i.e., $W_x^E(z,z') = P^{X,Z'}(z-x,z')$, combining (79), (80), and (74), we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3\frac{|\mathcal{X}|^t e^{\phi(t|X|Z'|P^{X,Z'})}}{L^t} \tag{84}$$

for $0 < \forall t < \frac{1}{2}$. That is, when $L = e^R$, taking the minimum concerning $0 < \forall t < \frac{1}{2}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3e^{-e_\phi(X|Z'|P^{X,Z'}|\log|\mathcal{X}|-R)}. \tag{85}$$

In the n-fold i.i.d. case, when $L = e^{nR}$, we obtain

$$\mathrm{E}_{\mathbf{X}} d_1(\Phi_{C_1,C_2(\mathbf{X})}|E) \le 3e^{-ne_\phi(X|Z'|P^{X,Z'}|\log|\mathcal{X}|-R)}. \tag{86}$$

When $\mathcal{X}$ is an n-dimensional vector space $\mathbb{F}_q^n$ over the finite field $\mathbb{F}_q$, the bound can be attained by the combination of linear code and the concatenation of Toeplitz matrix and the

identity $(\mathbf{X}, I)$ of the size $m \times (m-k)$[6]. Hence, if the error correcting code C_1 can be realizable, the whole process in the above code can be realizable.

Remark 1: In the additive case, due to (73), the exponent of the upper bound given in (83) is the same as that given by the previous paper [17]. However, the code given in [17] is constructed by completely random coding. However, the code given in this section is based on the ordinary linear code. For the security, it requires only the universal hash condition. So, our construction requires smaller complexity than that given in [17]. In the general additive case, our exponents (86) is strictly better than that given in [17], which is calculated in (77).

Next, we consider the relation with the other previous paper [6] in the general additive case. The protocol given in [6] is is quite similar to ours. However, as is shown in Lemma 8, except for the very special case, our exponent (86) is strictly better than that given in [6], which is calculated in (78). Remember that the exponent given in [6] is $\tilde{e}_\psi(R|W^E, p_{\mathrm{mix}})$, which is mentioned around (71).

IX. Secret key generation with public communication

Furthermore, the above result can be applied to secret key generation (distillation) with one-way public communication, in which, Alice, Bob, and Eve are assumed to have initial random variables $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $E \in \mathcal{E}$, respectively. The task for Alice and Bob is to share a common random variable almost independent of Eve's random variable E by using a public communication. For this purpose, we assume that Alice and Bob can perform local data processing in the both sides and Alice can send messages to Bob via public channel. That is, only one-way communication is allowed. We call such a combination of these operations a code and denote it by Φ.

The quality is evaluated by three quantities: the size of the final common random variable, the probability that their final variables coincide, and Eve's distinguishability $d_1(\Phi|E)$ of the final joint distribution between Alice and Eve.

In order to construct a protocol for this task, we assume that the set $\mathcal{A}$ has a module structure (any finite set can be regarded as a cyclic group). Then, the objective of secret key distillation can be realized by applying the code of a wire-tap channel as follows. First, Alice generates another uniform random variable X and sends the random variable $X' := X + A$. Then, the distribution of the random variables B, X' (E, X') accessible to Bob (Eve) can be regarded as the output distribution of the channel $x \mapsto W_x^B$ $(x \mapsto W_x^E)$. The channels W^B and W^E are given as follows.

$$W_x^B(x', b) = P^{A,B}(x' - x, b),\ W_x^E(x', e) = P^{A,E}(x' - x, e), \tag{87}$$

where $P^{AB}(a, b)$ $(P^{AE}(a, e))$ is the joint probability between Alice's initial random variable A and Bob's (Eve's) initial random variable B (E). Hence, the channel W^E is general additive.

Applying Theorem 5 to the uniform distribution P^A_{mix}, for any numbers M and L, due to (74), there exists a code Φ such that $|\Phi| = M$ and [1]

$$\epsilon_B(\Phi) \le 3 \min_{0\le s\le 1} (ML)^s |\mathcal{A}|^{-s} e^{\phi(-s|A|B|P^{A,B})} \tag{88}$$

$$d_1(\Phi|E) \le 9 \min_{0\le t\le \frac{1}{2}} \frac{|\mathcal{A}|^t e^{\phi(t|A|E|P^{A,E})}}{L^t}. \tag{89}$$

In particular, when the joint distribution between A and $B(E)$ is the n-fold independent and identical distribution (i.i.d.) of $P^{A,B}$ $(P^{A,E})$, respectively, the relation $\phi(t|A^n|E^n|(P^{A,E})^n) = n\phi(t|A|E|P^{A,E})$ hold. Thus, there exists a code Φ_n for any integers L_n, M_n, and any probability distribution p on $\mathcal{X}$ such that $|\Phi_n| = M_n$ and

$$\epsilon_B(\Phi) \le 3 \min_{0\le s\le 1} (M_n L_n)^s |\mathcal{A}|^{-ns} e^{n\phi(-s|A|B|P^{A,B})} \tag{90}$$

$$d_1(\Phi_n|E) \le 9 \min_{0\le t\le \frac{1}{2}} \frac{|\mathcal{A}|^{nt} e^{n\phi(t|A|E|P^{A,E})}}{L_n^t}. \tag{91}$$

Finally, we mention the relation with the previous paper [17]. Since the above discussion is an application of section VIII, the same comparison as Remark 1 is valid. Hence, our evaluation (91) is strictly better than that given in [17] except for the special case.

X. Discussion

We have derived the tight evaluation for exponent for the average of the L_1 norm distance between the generated random number and the uniform random number when universal$_2$ hash functions are applied and the key generation rate is less than the critical rate R_1. Using this evaluation, we have obtained an upper bound for Eve's distinguishability in secret key generation from a common random number without communication when a universal$_2$ hash functions are applied. Since our bound is based on the Rényi entropy of order $1+s$ for $s \in [0, 1]$, it can be regarded as an extension of Bennett et al [2]'s result with the Rényi entropy of order 2.

Applying this bound to the wire-tap channel, we obtain an upper bound for Eve's distinguishability, which yields an exponential upper bound. This exponent improves on the existing exponent [17]. Further, when the error correction code is given by a linear code and when the channel is additive or general additive, the privacy amplification is given by a concatenation of Toeplitz matrix and the identity matrix. This method can be applied to secret key distillation with public communication.

Acknowledgments

The author is grateful to Professors Ryutaroh Matsumoto and Takeshi Koshiba for a helpful comments. He is also grateful to the referee of the previous version for helpful comments.

[1] The previous paper [17, Section VI] derived upper bounds different from (88) and (90) while it treat the same protocol. The previous paper [17, Section VI] erroneously calculated $e^{\phi(-s|W^B, P_{\mathrm{mix},\mathcal{A}})}$ to $|\mathcal{A}|^{-s} e^{-sH_{\frac{1}{1+s}}(A|B|P^{A,B})}$. However, the correct calculation is $|\mathcal{A}|^{-s} e^{\phi(-s|A|B|P^{A,B})}$ as is shown in (74).

This research was partially supported by a MEXT Grant-in-Aid for Young Scientists (A) No. 20686026 and a MEXT Grant-in-Aid for Scientific Research (A) No. 23246071. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

APPENDIX A
PROOF OF THEOREM 2

First, for a fixed element $a \in \Omega$, we introduce the condition for a hash funcation $f_{\mathbf{X}}$:

Condition 5 (Condition $[a, \Omega]$):

$$f_{\mathbf{X}}(a) \neq f_{\mathbf{X}}(a') \text{ for } \forall a' (\neq a) \in \Omega.$$

Let $P[a, \Omega]$ be the probability that Condition $[a, \Omega]$ holds. Due to strongly universal$_2$ condition, it is evaluated by $P[a, \Omega] \ge 1 - \frac{|\Omega|}{M}$. When we denote the expectation concerning the hash funcations under Condition $[a, \Omega]$ by $\mathrm{E}_{\mathbf{X}|[a,\Omega]}$, the convexity of the function $x \mapsto |x|$ yields that

$$\begin{aligned}
&\mathrm{E}_{\mathbf{X}|[a,\Omega]} |P^A(a) + \sum_{a'(\neq a) \in f_{\mathbf{X}}^{-1}(a)} P^A(a') - \frac{1}{M}| \\
\ge & |P^A(a) + \mathrm{E}_{\mathbf{X}|[a,\Omega]} \sum_{a'(\neq a) \in f_{\mathbf{X}}^{-1}(a)} P^A(a') - \frac{1}{M}| \\
= & |P^A(a) + \frac{1}{M} \sum_{a'(\neq a) \in \mathcal{A} \setminus \Omega} P^A(a') - \frac{1}{M}| \\
= & |P^A(a) + \frac{1}{M}(1 - P^A(\Omega)) - \frac{1}{M}| \\
= & |P^A(a) - \frac{1}{M} P^A(\Omega)|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathrm{E}_{\mathbf{X}} d_1(P^{f_{\mathbf{X}}(A)}) \\
\ge & \sum_{a \in \Omega} P[a, \Omega] \mathrm{E}_{\mathbf{X}|[a,\Omega]} |P^A(a) + \sum_{a'(\neq a) \in f_{\mathbf{X}}^{-1}(a)} P^A(a') - \frac{1}{M}| \\
\ge & \sum_{a \in \Omega} (1 - \frac{|\Omega|}{M}) |P^A(a) - \frac{1}{M} P^A(\Omega)| \\
\ge & |\sum_{a \in \Omega} (1 - \frac{|\Omega|}{M})(P^A(a) - \frac{1}{M} P^A(\Omega))| \\
= & |(1 - \frac{|\Omega|}{M})(P^A(\Omega) - \frac{|\Omega|}{M} P^A(\Omega))| = (1 - \frac{|\Omega|}{M})^2 P^A(\Omega).
\end{aligned}$$

APPENDIX B
PROOF OF LEMMA 6

We choose $s(R)$ such that $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=s(R)} = H(P_{1+s(R)}) + D(P_{1+s(R)} \| P) = R$, where $P_{1+s}(a) := \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}}$. When Q satisfies $H(Q) + D(Q\|P) = R$,

$$\begin{aligned}
& D(Q\|P) - D(P_{1+s}\|P) \\
= & \sum_a Q(a)(\log Q(a) - \log P(a)) \\
& - \sum_a \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}} (\log \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}} - \log P(a)) \\
= & \sum_a Q(a)(\log Q(a) - \log \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}}) \\
& + \sum_a (Q(a) - \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}}) \\
& \cdot (\log \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}} - \log P(a)) \\
= & D(Q\|P_{1+s}) + s \sum_a (Q(a) - \frac{P(a)^{1+s}}{\sum_{a'} P(a')^{1+s}}) \log P(a) \\
= & D(Q\|P_{1+s}) \\
& + s(H(P_{1+s}) + D(P_{1+s}\|P) - H(Q) + D(Q\|P)) \\
= & D(Q\|P_{1+s}) \ge 0.
\end{aligned}$$

Hence,

$$\begin{aligned}
& \min_{Q: H(Q) + D(Q\|P) = R} H(Q) + 2D(Q\|P) - R \\
= & \min_{Q: H(Q) + D(Q\|P) = R} D(Q\|P) = D(P_{1+s(R)}\|P) \\
= & sH_{1+s}(A|P) - s(R) \frac{d(sH_{1+s}(A|P))}{ds}|_{s=s(R)} \\
= & sH_{1+s}(A|P) - s(R)R = \max_{0 \le s} sH_{1+s}(A|P) - sR.
\end{aligned}$$

The last equation follows from the concavity of $sH_{1+s}(A|P)$ concerning s.

Assume that $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} \le R$. Then, $s(R) \le 1$. When $R' \ge R$,

$$\begin{aligned}
& \min_{Q: H(Q) + D(Q\|P) = R'} H(Q) + 2D(Q\|P) - R \\
= & \max_{0 \le s} sH_{1+s}(A|P) - sR + R' - R \\
\ge & sH_{1+s(R)}(A|P) - s(R)R' + R' - R \\
\ge & sH_{1+s(R)}(A|P) - s(R)R \\
= & \max_{0 \le s} sH_{1+s}(A|P) - sR \\
= & \max_{0 \le s \le 1} sH_{1+s}(A|P) - sR,
\end{aligned}$$

which implies (51).

Assume that $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} > R$. When $R' \ge R$,

$$\begin{aligned}
& \min_{Q: H(Q) + D(Q\|P) = R'} H(Q) + 2D(Q\|P) - R \\
= & \max_{0 \le s} sH_{1+s}(A|P) - sR + R' - R \\
\ge & 1H_{1+1}(A|P) - R' + R' - R = H_2(A|P) - R.
\end{aligned}$$

Further, when $R' = \frac{d(sH_{1+s}(A|P))}{ds}|_{s=1}$,

$$\begin{aligned}
& \min_{Q: H(Q) + D(Q\|P) = R'} H(Q) + 2D(Q\|P) - R \\
= & H_{1+1}(A|P) - R' + R' - R = H_2(A|P) - R,
\end{aligned}$$

which implies (52).

Further, the concavity of $s \mapsto sH_{1+s}(A|P)$ and the condition $\frac{d(sH_{1+s}(A|P))}{ds}|_{s=1} > R$ imply that $\max_{0\le s\le 1} sH_{1+s}(A|P) - sR = H_2(A|P) - R$. Thus, we obtain (53).

REFERENCES

[1] L. Carter and M. Wegman, "Universal classes of hash functions," *J. Comput. Sys. Sci.*, vol. **18**(2), 143–154, 1979.

[2] C.H.Bennett, G. Brassard, C. Crepeau, and U.M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inform. Theory*, vol. **41**, 1915–1923, 1995.

[3] U. Maurer and S. Wolf, "Infromation-theoretic key agreement: From weak to strong secrecy for free," *Advances in Cryptology–EUROCRYPT 2000*, Lecture Notes in Computer Science, vol.1807, pp.351-368, Springer-Verlag (2000).

[4] R. Renner and S. Wolf, "Simple and Tight Bounds for Information Reconciliation and Privacy Amplification," *ASIACRYPT 2005*, Lecture Notes in Computer Science, Springer-Verlag, vol. 3788, pp. 199-216, 2005.

[5] R. Renner, "Security of Quantum Key Distribution," PhD thesis, Dipl. Phys. ETH, Switzerland, 2005. arXiv:quantph/0512258.

[6] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inform. Theory*, vol. **57**, No. 6, 3989–4001, 2011.

[7] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography part 1: Secret sharing," *IEEE Trans. Inform. Theory*, vol. **39**(4) 1121–1132, 1993.

[8] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. **39**, 733–742, 1993.

[9] J. Muramatsu, "Secret key agreement from correlated source outputs using low density parity check matrices," *IEICE Trans. Fundamentals*, E89-A(7): 2036-2046, 2006.

[10] J. Muramatsu, S. Miyake, "Construction of Codes for Wiretap Channel and Secret Key Agreement from Correlated Source Outputs by Using Sparse Matrices," arXiv:0903.4014. (2009).

[11] S. Watanabe, T. Saitou, R. Matsumoto, T. Uyematsu "Strongly Secure Privacy Amplification Cannot Be Obtained by Encoder of Slepian-Wolf Code," *Proceedings of the 2009 IEEE International Symposium on Information Theory*, Volume 2, Seoul, Korea, pp. 1298-1302 (2009) (arXiv:0906.2582)

[12] A. D. Wyner, "The wire-tap channel," *Bell. Sys. Tech. Jour.*, vol. **54**, 1355–1387, 1975.

[13] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. **24**(3) 339–348, 1979.

[14] I. Csiszár, "Almost Independence and Secrecy Capacity," *Problems of Information Transmission*, vol.**32**(1), pp.40-47, 1996.

[15] I. Devetak, "The private classical information capacity and quantum information capacity of a quantum channel," *IEEE Trans. Inform. Theory*, vol. **51**(1), 44–55, 2005.

[16] A. Winter, A. C. A. Nascimento, and H. Imai, "Commitment Capacity of Discrete Memoryless Channels," *Proc. 9th Cirencester Crypto and Coding Conf.*, LNCS 2989, pp 35-51, Springer, Berlin 2003; cs.CR/0304014 (2003)

[17] M. Hayashi, "General non-asymptotic and asymptotic formulas in channel resolvability and identification capacity and its application to wire-tap channel," *IEEE Trans. Inform. Theory*, vol. **52**(4), 1562–1575, 2006.

[18] H. Krawczyk, "LFSR-based hashing and authentication," *Advances in Cryptology — CRYPTO '94.*, Lecture Notes in Computer Science, vol. 839, Springer-Verlag, pp 129–139, 1994.

[19] I. Csiszár and J. Körner, *Information theory: Coding Theorem for Discrete Memoryless systems*, Academic Press, New York, (1981)

[20] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.

[21] P. Delsarte and P. Piret, "Algebraic constructions of Shannon codes for regular channels," *IEEE Trans. Inform. Theory*, vol.28, no.4, pp.593-599, 1982.

[22] R. Cannetti, "Universal composable security: a new paradigm for cryptographic protocols," *Proc. 42nd IEEE FOCS*, pp. 136-145, Oct. 2001.

[23] S. Watanabe, private communication, 2007. (This communication is written in [6, Appendix III])

[24] M. Hayashi, "Exponents of quantum fixed-length pure state source coding," *Physical Review A*, Vol.66, 032321 (2002).

[25] M. Hayashi, "Second-Order Asymptotics in Fixed-Length Source Coding and Intrinsic Randomness," *IEEE Trans. Inform. Theory*, **54**, 4619 - 4637 (2008).

[26] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: fundamental limits," *IEEE Trans. Inform. Theory*, **41**, 1322-1332 (1995).

[27] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, (Springer, 1997).

[28] L.P. Kuptsov, "Holder inequality", in Hazewinkel, Michiel, Encyclopaedia of Mathematics, Springer, (2001).

[29] T. Holenstein and R. Renner, "On the randomness of independent experiments," arXiv:cs/0608007 (2006)

[30] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby, "A Pseudorandom Generator from any One-way Function," SIAM J. Comput. 28, 1364 (1999)

[31] T.S. Han, "The reliability functions of the general source with fixed-length coding," *IEEE Trans. Inform. Theory*, **46**, 2117–2132, (2000).

[32] T. S. Han: *Information-Spectrum Methods in Information Theory*, (Springer-Verlag, New York, 2002) (Originally written in Japanese in 1998).

[33] M. R. Bloch, and J. N. Laneman, "Secrecy from Resolvability," submitted to *IEEE Trans. Inform. Theory*, arXiv:1105.5419 (2011).

[34] T.-H. Chou, V. Y. F. Tan, and S. C. Draper, "The Sender-Excited Secret Key Agreement Model: Capacity and Error Exponents," submitted to *IEEE Trans. Inform. Theory*, arXiv:1107.4148 (2011).